

受領書

平成10年 8月11日

特許庁長官

識別番号

100070150

氏名(名称)

伊東 忠彦

殿

提出日

平成10年 8月11日

以下の書類を受領しました。

| 項番 | 書類名 | 整理番号 | 受付番号 | 出願番号通知(事件の表示) |
|----|--------|-----------|-------------|-------------------------|
| 1 | 特許願 | ASIST9801 | 59800491955 | 特願平10-227278 |
| 2 | 出願審査請求 | — | 59800491962 | 特願平10. 8.11 (ASIST9801) |

以上

【書類名】 特許願

【整理番号】 ASIST9801

【提出日】 平成10年 8月11日

【あて先】 特許庁長官 伊佐山 建志 殿

【国際特許分類】 G06F 17/30

【発明の名称】 表形式データの検索・集計方法及び装置、並びに、検索・集計プログラムを記録した記録媒体

【請求項の数】 12

【発明者】

【住所又は居所】 神奈川県横浜市神奈川区松見町4丁目1101番地7
コートハウス菊名804号

【氏名】 古庄 晋二

【特許出願人】

【住所又は居所】 神奈川県横浜市神奈川区松見町4丁目1101番地7
コートハウス菊名804号

【氏名又は名称】 古庄 晋二

【代理人】

【識別番号】 100070150

【弁理士】

【氏名又は名称】 伊東 忠彦

【手数料の表示】

【予納台帳番号】 002989

【納付金額】 21000

【提出物件の目録】

【物件名】 明細書 1

【物件名】 図面 1

【物件名】 要約書 1

【書類名】 明細書

【発明の名称】 表形式データの検索・集計方法及び装置、並びに、検索・集計プログラムを記録した記録媒体

【特許請求の範囲】

【請求項1】 情報の項目に対応した項目値を含むレコードの配列として表される表形式データから、特定の項目及び特定のレコードに対応した項目値を得る方法において、

上記特定の項目に属する項目値に対応した項目値番号の順番に上記項目値が格納されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定する情報が格納された項目値番号指定情報配列とを記憶装置に保持し、

上記項目値番号指定情報配列から上記特定のレコードに対応した項目値番号を取得し、

上記値管理テーブルに格納された項目値の中から、上記取得された項目値番号に対応した項目値を得ることを特徴とする方法。

【請求項2】 上記項目値番号に対応した項目値を分類するための分類番号が上記項目値番号に対応して上記値管理テーブルに格納され、

上記項目値番号に対応した項目値を得る際に上記分類番号が参照されることを特徴とする請求項1記載の方法。

【請求項3】 情報の項目に対応した項目値を含むレコードの配列として表される表形式データから検索条件に適合する項目値を含むレコードを検索する方法において、

上記検索条件と関連した項目に属する項目値に対応した項目値番号の順番に上記項目値が格納されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定する情報が格納された項目値番号指定情報配列と、同じ項目値番号に対応した上記レコードの識別情報が上記項目値番号毎に別個に格納されたレコード識別情報配列とを含み、上記値管理テーブルは上記レコード識別情報配列の中で同じ項目値番号に対応したレコードの識別情報が格納されている場所を示すレコード識別情報指定情報が上記項目値番号毎に格納されている、個別項目情報を記憶装置に保持し、

上記値管理テーブルに格納された項目値の中で上記検索条件に適合する項目値と関連した項目値番号に対応して格納された上記レコード識別情報指定情報を用いて、上記レコード識別情報配列から上記検索条件に適合するレコードの識別情報を取り出すことを特徴とする方法。

【請求項4】 上記検索条件に適合するレコードの識別情報に基づいて上記検索条件に適合するレコードを指定する結果集合を記憶装置に保持し、

別の検索条件と関連した項目に関する別の個別項目情報を選択し、

上記別の個別項目情報の項目値番号指定情報配列から、上記結果集合で指定された上記検索条件に適合したレコードの識別情報に対応した項目値番号を取り出し、

上記別の個別項目情報について上記取り出された項目値番号と関連した項目値が上記別の検索条件と適合するか否かを判定し、

上記別の個別項目情報について上記項目値番号と関連した項目値が上記別の検索条件と適合する場合に、上記項目値番号に対応した上記レコードの識別情報を、上記別の検索条件に適合するレコードの識別情報として取り出すことを特徴とする請求項3記載の方法。

【請求項5】 情報の項目に対応した項目値を含むレコードの配列として表される表形式データを集計する方法において、

n が1以上の整数を表す場合に、集計に利用される n 個の項目について、上記項目に属する項目値に対応した項目値番号の順番に上記項目値が格納されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定する情報が格納された項目値番号指定情報配列とにより表された n 個の個別項目情報を記憶装置に保持し、

i が $1 \leq i \leq n$ なる整数を表すとき、 i 番目の個別情報項目について、上記項目値番号の総数が N_i で表され、 k_i が $0 \leq k_i \leq N_i - 1$ なる整数を表し、 M が1以上の整数を表し、 m が $1 \leq m \leq M$ なる整数であるとき、 $N_1 \times N_2 \times \dots \times N_1 \times \dots \times N_n$ のサイズを持つ n 次元の M 個のデータ空間の各要素 $P_m(k_1, k_2, \dots, k_1, \dots, k_n)$ を初期化し、

上記 n 個の個別情報項目について、 j が $0 \leq j \leq (\text{レコードの総数} - 1)$ なる

整数を表すとき、各項目値番号指定情報配列の j 番目に格納された夫々の項目値番号を取り出し、 i 番目の個別情報項目から取り出された項目値番号が q_i で表されるとき、上記データ空間の要素 $P_m (q_1, q_2, \dots, q_i, \dots, q_n)$ を特定し、

上記特定された要素 $P_m (q_1, q_2, \dots, q_i, \dots, q_n)$ の値を処理する段階を含むことを特徴とする方法。

【請求項6】 情報の項目に対応した項目値を含むレコードの配列として表される表形式データを集計する方法において、

n が1以上の整数を表し、集計に利用される n 個の項目について、上記項目に属する項目値に対応した項目値番号の順番に上記項目値及び上記項目値の分類番号が格納されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定する情報が格納された項目値番号指定情報配列とにより表された n 個の個別項目情報を記憶装置に保持し、

i が $1 \leq i \leq n$ なる整数を表すとき、 i 番目の個別情報項目について、上記項目値番号又は上記分類番号の何れか一方の番号の総数が N_i で表され、 k_i が $0 \leq k_i \leq N_i - 1$ なる整数を表し、 M が1以上の整数を表し、 m が $1 \leq m \leq M$ なる整数であるとき、 $N_1 \times N_2 \times \dots \times N_i \times \dots \times N_n$ のサイズを持つ n 次元の M 個のデータ空間の各要素 $P_m (k_1, k_2, \dots, k_i, \dots, k_n)$ を初期化し、

上記 n 個の個別情報項目について、 j が $0 \leq j \leq (\text{レコード総数} - 1)$ なる整数を表すとき、各項目値番号指定情報配列の j 番目に格納された夫々の項目値番号を取り出し、 i 番目の個別情報項目から取り出された項目値番号又は上記 i 番目の個別情報項目の値管理テーブルに該項目値番号と対応して格納された分類番号が q_i で表されるとき、上記データ空間の要素 $P_m (q_1, q_2, \dots, q_i, \dots, q_n)$ を特定し、

上記特定された要素 $P_m (q_1, q_2, \dots, q_i, \dots, q_n)$ の値を処理する段階を含むことを特徴とする方法。

【請求項7】 $M=1$ であり、

上記特定された要素 P_m の値を処理する段階は、上記要素 P_m の現在の値に1

を加える請求項5又は6記載の方法。

【請求項8】 上記特定された要素 P_m の値を処理する段階は、
M個の要素 P_m の中の少なくとも1個の要素 P_m に対して、
記憶装置に保持された別の個別項目情報について項目値番号指定情報配列のj番目に格納された項目値番号を取得し、

上記別の個別項目情報の値管理テーブルに格納された項目値の中から、上記取得された項目値番号に対応した項目値を取得し、

上記要素 P_m の現在の値と上記取得された項目値を組み合わせることで上記要素 P_m の値を更新する請求項5又は6記載の方法。

【請求項9】 上記項目値番号を指定する情報は項目値番号自体である請求項1乃至8のうちいずれか1項記載の方法。

【請求項10】 上記項目値番号を指定する情報は、各項目値番号毎に1ビット以上が割り当てられ、指定の有無が設定される2進値である請求項1乃至8のうちいずれか1項記載の方法。

【請求項11】 情報の項目に対応した項目値を含むレコードの配列として表される表形式データを検索・集計する装置において、

上記項目に属する項目値に対応した項目値番号の順番に上記項目値が格納されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定する情報が格納された項目値番号指定情報配列とを保持する記憶装置と、

上記記憶装置に保持された上記項目値番号指定情報配列から上記特定のレコードに対応した項目値番号を取得する手段と、

上記記憶装置に保持された上記値管理テーブルに格納された項目値の中から、上記取得された項目値番号に対応した項目値を得る手段とを有することを特徴とする装置。

【請求項12】 情報の項目に対応した項目値を含むレコードの配列として表される表形式データを検索・集計するプログラムを記録した記録媒体において、

上記項目に属する項目値に対応した項目値番号の順番に上記項目値が格納されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定する情報

が格納された項目値番号指定情報配列とを記憶装置に保持させるプロセスと、

上記記憶装置に保持された上記項目値番号指定情報配列から上記特定のレコードに対応した項目値番号を取得させるプロセスと、

上記記憶装置に保持された上記値管理テーブルに格納された項目値の中から、上記取得された項目値番号に対応した項目値を得させるプロセスとを有することを特徴とするプログラムを記録した記録媒体。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】

本発明は、コンピュータのような情報処理装置を用いて大量のデータを処理するデータ処理方法及びデータ処理装置に係わり、特に、表形式のデータを検索・集計する方法及び装置に関する。

【0002】

【従来の技術】

従来、大量の情報を蓄積し、蓄積された情報を検索、集計するデータ処理が行われている。これらのデータ処理は、例えば、CPUと、メモリと、周辺機器インタフェースと、ハードディスクのような補助記憶装置と、ディスプレイ及びプリンタのような表示装置と、キーボード及びマウスのような入力装置と、電源ユニットとがバスを介して接続された周知のコンピュータシステムで使用され、特に、市場で容易に入手可能なコンピュータシステムで動作可能なソフトウェアとして提供される。上記の検索・集計等のデータ処理を行うため、特に、大量のデータを蓄積する各種データベースが知られている。大量のデータの中でも、特に、表形式で表現され得るデータを処理したいという要求は強い。図1は、処理されるべきデータを表形式で表現した例を示す図である。図1には、たとえば、100万人の多数の人に関して各人の性別、年齢及び職業のデータを蓄積したデータが示されている。図1において、表の横方向の行、即ち、所謂レコードは、レコード番号と、レコード番号に対応した性別、年齢及び職業の項目とからなる。表の縦方向の列は、レコード番号と、性別の項目と、年齢の項目と、職業の項目とからなる。レコード番号0の人は、性別が女性、年齢が18歳、職業がプログ

ラマであることを表す。以下の説明では、女性、18歳、プログラマ等のように各項目に設定されたデータを項目値と称する。また、以下の説明では、特に断らない限り、大量のデータの具体的な例として、図1に示された100万レコードからなる表形式データを使用する。

【0003】

大量のデータを効率よく検索、集計できるか否かは、大量のデータを格納する形式に依存する。従来、一般的な格納技術として図2の(A)及び(B)に夫々示された所謂「行単位」格納技術及び「項目単位」格納技術が知られている。

図2には、記憶装置、たとえば、ハードディスク上のデータ格納イメージを示している。図2の(A)の行単位格納技術の場合、各レコード番号毎に構成された性別、年齢及び職業の項目値の組がレコード番号順に、論理的アドレスが増加する順番でディスク上に格納されている。一方、図2の(B)の項目単位格納技術の場合、各項目毎にレコード番号順に項目値が論理的アドレスが増加する方向にディスク上に格納されている。即ち、図2の(B)の例では、性別に関してレコード番号0から999999までの項目値が順番に並べられ、次に、年齢に関してレコード番号順に項目値が並べられ、以下同様である。

【0004】

上記の従来技術の場合、全てのレコード番号の全ての項目に対する項目値がそのまま(レコード番号の1次元と、それ以外の項目値の1次元とからなる)2次元のデータ構造に格納されている。以下、このようなデータ構造を特に「データ表」と称する。従来技術の場合、蓄積されたデータを検索、集計する際には、このデータ表をアクセスすることにより行われている。

【0005】

また、項目の値をそのまま項目値として格納する方法の他に、値をコード変換して、項目値としてコードを格納する方法も知られている。たとえば、性別に関して値「男性」を0に、値「女性」を1にコード変換し、「男性」又は「女性」の代わりに0又は1の値が項目値として格納される。この場合でも、コード変換されたコードが項目値としてデータ表に格納されている点に変わりはない。

【0006】

上記従来技術のデータ表タイプのデータ構造を用いて格納された大量のデータを検索、集計する場合、このようなデータ表をアクセスするためのアクセス時間に起因して検索、集計の処理時間が長くなるという問題点がある。

また、データ表には以下の3点の本質的な欠点がある。

欠点1：データ表はサイズが巨大化しやすく、しかも例えば項目毎等に（物理的に）分割することが困難である。例えば、性別が「男性」のレコードを抽出するとき、年齢や職業の情報は必要とされないので、性別という項目だけからなる表に分割してアクセスできれば効率的である。図2の（B）に示された項目単位格納技術の場合、項目毎への分割はしやすいが、大量のデータを扱う際にはデータ表のサイズが巨大化することは同じで、實際上、集計・検索のためにデータ表をメモリなどの高速な記憶装置上に展開することは困難である。

【0007】

欠点2：データ表は、各項目値を同時にソートした形で保持できない。例えば、図2の（A）及び（B）に示した従来技術の場合、項目値は、性別に関して「女性、男性、女性、．．．、女性」のようにレコード番号順に並べられている。しかし、検索・集計処理を行う場合、一般に、「女性、女性、女性、男性、．．．、男性」のように並べられている方が便宜である。ところが、データ表では、項目値は所定の配列順序、即ち、レコード番号順に並べられているので、特定の項目に関して項目値をソートすることは許容されない。そのため、従来技術の場合、検索・集計に都合の良い項目値の並べ方を選択することができない。

【0008】

欠点3：データ表には、同一値が何度も出現する。例えば、図2の（A）及び（B）に記載された従来のデータ表の場合、性別が『「男性」か「男」』であるレコード、即ち、レコード番号を抽出する操作の際に、「男性」という項目値が何度も出現することに起因して、「男性」という項目値と比較条件である『「男性」か「男」』との照合を何度も行わねばならないことになる。本来的には、同一値に対して、それが該当するか否かは、1回判定すれば良いはずである。

【0009】

【発明が解決しようとする課題】

本発明の目的は、大量のデータの検索・集計のスピードを大幅に向上させるため、従来のデータ表の機能を有すると共に、データ表に基づくデータ構造の上記問題点が解決されたデータ管理機構を設けることにより、表形式データを検索・集計する方法及びその方法を実施する装置を提供することである。

【0010】

【課題を解決するための手段】

上記目的を達成するため、本発明の表形式データを検索・集計する方法及び装置は、通常のコンピュータシステムで使用可能な新規のデータ管理機構を提案する。本発明によるデータ管理機構は原則として値管理テーブルと、値管理テーブルへのポインタ配列とを有する。

【0011】

図3は、本発明の原理を説明する図であり、値管理テーブル10と値管理テーブルへのポインタ配列20とが示されている。値管理テーブル10とは、表形式データの各項目に対し、その項目に属する項目値が順序付け（整数化）された項目値番号の順番に、上記項目値番号に対応した項目値11と、上記項目値に関連した分類番号12とが格納されたテーブルである。値管理テーブルへのポインタ配列20とは、表形式データのある列（即ち、項目）の項目値番号、即ち値管理テーブル10へのポインタが表形式データのレコード番号順に格納された配列である。

【0012】

値管理テーブルへのポインタ配列20と値管理テーブル10とを組み合わせることにより、あるレコード番号が与えられたとき、所定の項目に関する値管理テーブルへのポインタ配列20からそのレコード番号に対応して格納された項目値番号を取り出し、次に、値管理テーブル10内でその項目値番号に対応して格納された項目値を取り出すことにより、レコード番号から項目値を得ることができる。したがって、従来のデータ表と同様に、レコード番号（行）と項目（列）という座標を用いて全てのデータ（項目値）を参照することができる。

【0013】

このように表形式データの項目の中の一つの項目に対して生成された値管理テ

ーブルと値管理テーブルへのポインタ配列とを含む本発明によるデータ管理機構は、以下の説明では、特に情報ブロックと呼ぶ場合がある。

従来のデータ表は、レコードに対応した行と、項目に対応した列とからなる座標を用いて全てのデータを一体的に管理しているのに対し、本発明による情報ブロックは、表形式の列、即ち、項目毎にデータを完全に分離している点に特徴がある。このように、本発明によれば、大量のデータが項目毎に分離されているので、検索・集計に必要な項目に関するデータのみをメモリ等の高速な記憶装置上に取り込むことが可能であり、その結果としてデータへのアクセス時間が短縮されるので、検索・集計の処理が高速化され、項目数が非常に多いデータの場合でも、パフォーマンスを低下させることなく取り扱えるようになる。

【0014】

また、本発明による情報ブロックの場合、項目値は値管理テーブルに格納され、値が存在する位置を示すレコード番号は値管理テーブルへのポインタ配列に関連付けられているので、項目値がレコード番号順に並べられている必要がない。したがって、検索・集計に適するようにデータを項目値に関してソートすることができるようになる。これにより、目的の値と一致する項目値がデータ中に存在するか否かの判定が高速に行えるようになる。その上、項目値は項目値番号と対応しているので、項目値が長いデータや文字列等であっても整数として取り扱うことができる。

【0015】

さらに、本発明によれば、値管理テーブル10の全ての項目値番号は異なる項目値と対応しているので、特定の値を有する項目値を含むレコードを抽出する場合に必要とされる特定の値と項目値との比較の回数は、最大で項目値の種類の数、即ち、項目値番号の個数であり、比較演算の回数が著しく低減され、検索・集計の高速化が図られる。その際には、ある項目値が該当するかどうかを調べた結果を格納する場所が必要であるが、例えば分類番号12をその格納場所として使用することができる。

【0016】

図4には、項目値11と、分類番号12と、存在数14とからなる値管理テー

ブル10を含む本発明による情報ブロックが示されている。存在数14には、ある項目に関する各項目値が全データ中に何個ずつ存在するかを示す数、換言すれば、所定の項目値を有するレコードの個数が格納される。このような存在数14を値管理テーブル10に準備しておくことにより、検索・集計の際に必要とされる「どのようなデータが(いくつ)存在するか」という情報を直ちに得ることができるようになり、検索・集計の高速化が図れる。

【0017】

図5には、値管理テーブル10と、値管理テーブルへのポインタ配列20と、レコードへのポインタ配列30とからなる情報ブロックが示されている。レコードへのポインタ配列30とは、項目値番号、即ち、項目値毎に、その項目値を有するレコードへのポインタ(レコード番号に対応する)が格納された配列である。各項目値毎にレコードへのポインタ配列30に格納されるポインタの個数は、値管理テーブル10の存在数14に格納された数と一致する。また、レコードへのポインタ配列30内で、各項目値毎のポインタの群の先頭のアドレスを指定する開始位置13を値管理テーブル10に設けてもよい。このようなレコードへのポインタ配列30を情報ブロックに設けることにより、所定の項目に関して、特定の項目値を有するレコードの集合を高速に抽出できるようになる。値管理テーブル10にレコードへのポインタ配列30に格納されるポインタの存在数14及び開始位置13が設定されているので、集計時に利用可能な値、個数が情報ブロックにそのまま存在している点が有利である。

【0018】

次に、本発明による表形式データの検索・集計方法について説明する。尚、以下の説明では、個別項目情報は上記の「情報ブロック」を表し、項目値番号指定情報配列は上記の「値管理テーブルへのポインタ配列」を表し、レコード識別情報配列は上記の「レコードへのポインタ配列」を表す。

本発明による情報の項目に対応した項目値を含むレコードの配列として表される表形式データから、特定の項目及び特定のレコードに対応した項目値を得る方法は、

上記特定の項目に属する項目値に対応した項目値番号の順番に上記項目値が格

納されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定する情報が格納された項目値番号指定情報配列とを記憶装置に保持し、

上記項目値番号指定情報配列から上記特定のレコードに対応した項目値番号を取得し、

上記値管理テーブルに格納された項目値の中から、上記取得された項目値番号に対応した項目値を得る段階を含む。

【0019】

また、本発明による項目値を得る方法によれば、上記項目値番号に対応した項目値を分類するための分類番号が上記項目値番号に対応して上記値管理テーブルに格納され、上記項目値番号に対応した項目値を得る際に上記分類番号が参照される。

本発明による情報の項目に対応した項目値を含むレコードの配列として表される表形式データから検索条件に適合する項目値を含むレコードを検索する単項目検索方法は、

上記検索条件と関連した項目に属する項目値に対応した項目値番号の順番に上記項目値が格納されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定する情報が格納された項目値番号指定情報配列と、同じ項目値番号に対応した上記レコードの識別情報が上記項目値番号毎に別個に格納されたレコード識別情報配列とを含み、上記値管理テーブルは上記レコード識別情報配列の中で同じ項目値番号に対応したレコードの識別情報が格納されている場所を示すレコード識別情報指定情報が上記項目値番号毎に格納されている、個別項目情報を記憶装置に保持し、

上記値管理テーブルに格納された項目値の中で上記検索条件に適合する項目値と関連した項目値番号に対応して格納された上記レコード識別情報指定情報を用いて、上記レコード識別情報配列から上記検索条件に適合するレコードの識別情報を取り出す段階を含む。

【0020】

また、本発明による複数項目検索方法は、

本発明の単項目検索方法に従って得られた上記検索条件に適合するレコードの

識別情報に基づいて上記検索条件に適合するレコードを指定する結果集合を記憶装置に保持し、

別の検索条件と関連した項目に関する別の個別項目情報を選択し、
上記別の個別項目情報の項目値番号指定情報配列から、上記結果集合で指定された上記検索条件に適合したレコードの識別情報に対応した項目値番号を取り出し、

上記別の個別項目情報について上記取り出された項目値番号と関連した項目値が上記別の検索条件と適合するか否かを判定し、

上記別の個別項目情報について上記項目値番号と関連した項目値が上記別の検索条件と適合する場合に、上記項目値番号に対応した上記レコードの識別情報を上記別の検索条件に適合するレコードの識別情報として取り出す段階を含む。

【0021】

本発明による情報の項目に対応した項目値を含むレコードの配列として表される表形式データを項目値毎に集計する方法は、

n が1以上の整数を表す場合に、集計に利用される n 個の項目について、上記項目に属する項目値に対応した項目値番号の順番に上記項目値が格納されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定する情報が格納された項目値番号指定情報配列とにより表された n 個の個別項目情報を記憶装置に保持し、

i が $1 \leq i \leq n$ なる整数を表すとき、 i 番目の個別情報項目について、上記項目値番号の総数が N_i で表され、 k_i が $0 \leq k_i \leq N_i - 1$ なる整数を表し、 M が1以上の整数を表し、 m が $1 \leq m \leq M$ なる整数であるとき、 $N_1 \times N_2 \times \dots \times N_1 \times \dots \times N_n$ のサイズを持つ、 n 次元の、 M 個のデータ空間の各要素 $P_m(k_1, k_2, \dots, k_i, \dots, k_n)$ を初期化し、

上記 n 個の個別情報項目について、 j が $0 \leq j \leq (\text{レコードの総数} - 1)$ なる整数を表すとき、各項目値番号指定情報配列の j 番目に格納された夫々の項目値番号を取り出し、 i 番目の個別情報項目から取り出された項目値番号が q_i で表されるとき、上記データ空間の要素 $P_m(q_1, q_2, \dots, q_i, \dots, q_n)$ を特定し、

上記特定された要素 $P_m (q_1, q_2, \dots, q_i, \dots, q_n)$ の値を処理する段階を含む。

【0022】

また、本発明による情報の項目に対応した項目値を含むレコードの配列として表される表形式データを項目値の分類毎に集計する方法は、

n が 1 以上の整数を表し、集計に利用される n 個の項目について、上記項目に属する項目値に対応した項目値番号の順番に上記項目値及び上記項目値の分類番号が格納されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定する情報が格納された項目値番号指定情報配列とにより表された n 個の個別項目情報を記憶装置に保持し、

i が $1 \leq i \leq n$ なる整数を表すとき、 i 番目の個別情報項目について、上記項目値番号又は上記分類番号の何れか一方の番号の総数が N_i で表され、 k_i が $0 \leq k_i \leq N_i - 1$ なる整数を表し、 M が 1 以上の整数を表し、 m が $1 \leq m \leq M$ なる整数であるとき、 $N_1 \times N_2 \times \dots \times N_i \times \dots \times N_n$ のサイズを持つ、 n 次元の、 M 個のデータ空間の各要素 $P_m (k_1, k_2, \dots, k_i, \dots, k_n)$ を初期化し、

上記 n 個の個別情報項目について、 j が $0 \leq j \leq (\text{レコード総数} - 1)$ なる整数を表すとき、各項目値番号指定情報配列の j 番目に格納された夫々の項目値番号を取り出し、 i 番目の個別情報項目から取り出された項目値番号又は上記 i 番目の個別情報項目の値管理テーブルに該項目値番号と対応して格納された分類番号が q_i で表されるとき、上記データ空間の要素 $P_m (q_1, q_2, \dots, q_i, \dots, q_n)$ を特定し、

上記特定された要素 $P_m (q_1, q_2, \dots, q_i, \dots, q_n)$ の値を処理する段階を含む。

【0023】

特に、本発明による個数を集計する方法によれば、 $M = 1$ であり、上記特定された要素 P_m の値を処理する段階は、上記要素 P_m の現在の値に 1 を加える。

また、本発明による統計的な量を集計する方法によれば、上記特定された要素 P_m の値を処理する段階は、 M 個の要素 P_m の中の少なくとも 1 個の要素 P_m に

対して、

記憶装置に保持された別の個別項目情報について項目値番号指定情報配列の j 番目に格納された項目値番号を取得し、

上記別の個別項目情報の値管理テーブルに格納された項目値の中から、上記取得された項目値番号に対応した項目値を取得し、

上記要素 P_m の現在の値と上記取得された項目値を組み合わせることで上記要素 P_m の値を更新する。

【0024】

本発明によれば上記項目値番号を指定する情報は項目値番号自体でもよい。

或いは、あるレコードの一つの項目に対し複数の項目値が割り当てられようないわゆるマルチアンサ項目を実現するため、本発明によれば、上記項目値番号を指定する情報は、各項目値番号毎に1ビット以上が割り当てられ、指定の有無が設定される2進値でも構わない。

【0025】

また、本発明による情報の項目に対応した項目値を含むレコードの配列として表される表形式データを検索・集計する装置は、

上記項目に属する項目値に対応した項目値番号の順番に上記項目値が格納されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定する情報が格納された項目値番号指定情報配列とを保持する記憶装置と、

上記記憶装置に保持された上記項目値番号指定情報配列から上記特定のレコードに対応した項目値番号を取得する手段と、

上記記憶装置に保持された上記値管理テーブルに格納された項目値の中から、上記取得された項目値番号に対応した項目値を得る手段とを有することを特徴とする。

【0026】

また、本発明による情報の項目に対応した項目値を含むレコードの配列として表される表形式データを検索・集計するプログラムを記録した記録媒体は、

上記項目に属する項目値に対応した項目値番号の順番に上記項目値が格納されている値管理テーブルと、上記レコードの順番に上記項目値番号を指定する情報

が格納された項目値番号指定情報配列とを記憶装置に保持させるプロセスと、

上記記憶装置に保持された上記項目値番号指定情報配列から上記特定のレコードに対応した項目値番号を取得させるプロセスと、

上記記憶装置に保持された上記値管理テーブルに格納された項目値の中から、上記取得された項目値番号に対応した項目値を得させるプロセスとを有することを特徴とするプログラムを記録した記録媒体である。

【0027】

【発明の実施の形態】

以下、本発明がより良く理解できるように図1に示した表形式データをデータの例として使用して、本発明の検索方法及び集計方法の種々の実施例を詳細に説明する。図1の例に示されるデータは、「性別」、「年齢」及び「職業」の項目を含むので、情報ブロックとして、図6乃至8に夫々示されているように、「性別」に関する情報ブロック、「年齢」に関する情報ブロック及び「職業」に関する情報ブロックが得られる。以下の実施例の説明では、これらの情報ブロックが得られた状況を想定している。尚、情報ブロックを構築する手法の一例は後述するが、本発明は情報ブロックを構築する方法によって制限されないことに注意する必要がある。

【0028】

「年齢」の項目値が16歳又は19歳であるレコードを検索する場合に関して、本発明の第1の実施例による単一項目内の検索方法を説明する。図9は、単一項目内の検索方法の動作フローチャートである。

最初に、表形式データに関する情報ブロックの中から、図7に示された「年齢」に関する情報ブロックを特定情報ブロックとして選択する（ステップ100）。

【0029】

次に、特定情報ブロックの値管理テーブルの中で項目値が上記の検索条件である16又は19に合致する行の分類番号を1にセットし、それ以外の行の分類番号を0にセットする（ステップ102）。図10に示されるように本例の場合、項目値番号0及び項目値番号3に対応する行の分類番号が1にセットされる。

続いて、分類番号が1にセットされている行（即ち、項目値番号0及び3）に対応した開始位置及び存在数をポインタ取り出し情報として取得する（ステップ104）。本例の場合、項目値番号0の開始位置は0であり、存在数は45898であり、項目値番号3の開始位置は238137であり、存在数は189653である。

【0030】

レコードへのポインタ配列の中で、上記の開始位置から存在数で指定された個数分のポインタを取り出すことにより、上記検索条件に合致したレコードへのポインタを表すレコード番号が取り出される（ステップ106）。本例の場合、図56に示されているように、項目値番号0に対応したレコードへのポインタは、レコードへのポインタ配列の開始位置0、即ち、先頭から45898個目までの場所に格納され、項目値番号3に対応したレコードへのポインタは、レコードへのポインタ配列の2383137番目から189653個分の場所に格納されていることが分かる。例えば、図1の表形式データを参照するに、最後のレコード番号999999のレコードに対応する「年齢」は16歳であるので、図11に示される如く、レコードポインタへの配列内で項目番号0、即ち、16歳の「年齢」に対応して格納されたポインタの中で最後のポインタは、999999である。

【0031】

最後に、後の処理で利用するため、取り出されたレコード番号の配列を結果集合として作成、保持する（ステップ108）。

以下、本発明の第2の実施例による複数項目のAND検索方法を説明する。本例では、「年齢」が16歳又は19歳であるという第1の検索条件と、「職業」が学生であるという第2の検索条件を満たすレコードの集合を得る場合を考える。図12は、複数項目のAND検索方法の動作フローチャートである。

【0032】

既に説明したように、第1の項目である「年齢」に関する情報ブロックである第1の特定情報ブロックについて、「年齢」が16歳又は19歳であるレコードの結果集合は本発明の第1の実施例に従って得られる（ステップ120）。

次に、図8に示されている第2の項目である「職業」に関する情報ブロックである第2の特定情報ブロックとして選択する（ステップ122）。

【0033】

次に、第2の特定情報ブロックの値管理テーブルの中で、項目値が上記の検索条件である学生に合致する行の分類番号を1にセットし、それ以外の行の分類番号を0にセットする（ステップ124）。図13に示されるように本例の場合、項目値番号0に対応する行の分類番号が1にセットされ、それ以外の行は0にセットされる。

【0034】

次に、第1の検索条件に関する結果集合から、レコードへのポインタを表すレコード番号を順次取り出す（ステップ126）。本例の場合、図14に示されるように、例えば、レコード番号999999が取り出される。

次に、第2の特定情報ブロックに関して、値管理テーブルへのポインタ配列から上記第1の検索条件について得られたレコード番号に対応した項目値番号を取り出す（ステップ128）。本例の場合、図14に示されるように、例えば、レコード番号999999に対応した項目値番号0が取り出される。

【0035】

次に、第2の特定情報ブロックに関して取り出された項目値番号に対応する分類番号が1にセットされているか否かを判定する（ステップ130）。本例の場合、図14に示されているように、例えば、項目値番号0に対応する分類番号は1にセットされていることがわかる。

分類番号が1にセットされている場合、値管理テーブルへのポインタ配列内で、分類番号が1にセットされている当該項目値番号を指すポインタが格納されている場所に対応したレコードへのポインタ、例えば、レコード番号を最終的な結果集合に追加する（ステップ132）。本例では、図14に示されているように、例えば、レコード番号999999が最終的な結果集合に追加される。

【0036】

分類番号が0である場合には、最終的な結果集合は更新されない。

尚、当業者であれば容易に分かるように、上記本発明の第2の実施例による複

数項目AND検索方法は、AND以外の検索、例えば、複数項目OR検索方法に変形することが可能である。即ち、最初に、第1の検索条件について結果集合を取得し、第2の検索条件について分類番号をセットする。次に、第1の検索条件の結果集合に格納されたレコード番号をスキップしながら、第2の特定情報ブロックについての値管理テーブルへのポインタ配列を順次走査する。その際、第1の検索条件の結果集合に格納されたレコード番号と、第2の情報ブロックについて分類番号がセットされた項目値番号に属するレコード番号とを併せて第2の結果集合として出力することができる。

【0037】

図15及び16は本発明の第3の実施例によるビットフラグを用いた複数項目論理演算検索方法の説明図であり、上記本発明の第2の実施例による検索と同一の検索条件で検索を行う場合を示している。ビットフラグを用いた複数項目論理演算検索とは、検索条件が項目毎の検索条件の間の論理演算によって表されるような検索を意味する。この場合、単一項目の検索により得られる結果集合を本発明の第1の実施例の如く、レコード番号の配列により構成するのではなく、結果集合をビットフラグの形式で構成する方が有利である。即ち、全レコードに対し1ビットずつを割り当て、各レコードが検索条件に適合しているか否かをビットの値1又は0によって示すように結果集合を構成する。結果集合をこのように構成することにより、各項目毎の結果集合のサイズが表形式データのレコード数と対応するので、各項目毎の結果集合のサイズが一致し、その結果として、結果集合間の論理演算、例えば、AND, OR, XORを簡単に行うことができる。

【0038】

本例では、図15に示された結果集合Aと図16に示された結果集合BをAND条件で結合することにより所望の検索結果集合がビットフラグ形式で得られる。尚、得られたビットフラグ形式の検索結果集合は、レコードへのポインタ配列形式の結果集合に変換することにより、上記本発明の第2の実施例による複数項目検索方法と組み合わせて用いることができる。

【0039】

以下、本発明による各種集計の作成方法の実施例を説明する。本発明の第4の

実施例による集計方法は、特定の項目に関して特定の項目値を有するレコード数をカウントする。本発明の第4の実施例では、特に、図1の表形式データに関して、項目「性別」について「男性」という項目値及び「女性」という項目値を有するレコードの数を夫々カウントする場合を考える。既に、図6に示しているように、本発明の好ましい実施例によれば、「性別」に関する情報ブロックには、項目値「男性」を含むレコードの存在数は632564個であり、項目値「女性」を含むレコードの存在数は367436個であることが格納されているので、単純なレコード数の集計は情報ブロックを参照することによって直ちに得ることができる。

【0040】

また、単一項目の検索方法について説明した分類番号を、上記本発明の第4の実施例による集計方法と組み合わせることにより更に複雑な条件の場合にもレコード数のカウントを簡単に行うことができる。例えば、本発明の第1の実施例で説明した「年齢」の項目値が16歳又は19歳であるレコードを検索する単一項目検索方法において、分類番号が1にセットされた項目値番号に対応した存在数を合計することにより、検索条件に適合したレコード数を集計することが可能である。このように、分類番号を使用することにより、値管理テーブルのサイズが大きい場合、或いは、複雑な条件が与えられた場合に、存在数を効率的に求めることができるようになる。

【0041】

次に、本発明の第5の実施例による男性の平均年齢を集計する方法を説明する。平均年齢は、(男性の「年齢」の総和) / (男性の存在数) により求めることができ、また、男性の存在数は、上記の本発明の第4の実施例で説明した集計方法により得ることができる。従って、本実施例は、男性の「年齢」の総和を求める問題に帰着される。図17は本発明の第5の実施例の動作フローチャートである。

【0042】

最初に、図6に示された「性別」に関する情報ブロックを第1の情報ブロックとして選択し(ステップ140)、第1の特定情報ブロックの値管理テーブルの

中で、項目値「男性」に対応した項目値番号0を検出する(ステップ142)。次に、項目値番号0に対応した存在数が632564であることから男性の総数が632564人であることを判定し(ステップ144)、また、項目値番号0に対応した開始位置が0であることから、性別が男性であるレコードへのポインタは、レコードへのポインタ配列の先頭から632564個目までの場所に格納されていることを判定し、このレコードへのポインタの一覧、即ち、レコード番号の配列を結果集合として保持する(ステップ146)。

【0043】

次に、図7に示された「年齢」に関する情報ブロックを第2の特定情報ブロックとして選択し(ステップ148)、第2の特定情報ブロックの値管理テーブルへのポインタ配列から、第1の特定情報ブロックに関する結果集合で指定されたレコードに対応した項目値番号を取り出し(ステップ150)、取り出された項目値番号と関連した項目値、即ち、年齢を取り出す(ステップ152)。最後に、取り出された年齢を順次加算して年齢の総和を求め(ステップ154)、上記結果集合で指定されたレコードを全て処理するまでステップ150、152、154を繰り返す(ステップ156)。得られた年齢の総和を存在数で除算することにより年齢の平均が求められる(ステップ158)。

【0044】

本発明の第6の実施例による学生の男性の平均年齢及び学生の女性の平均年齢を求める集計方法を説明する。図18は本発明の第6の実施例を概念的に説明する図であり、図19は本発明の第6の実施例の動作フローチャートである。

本実施例の集計は、最初に、第1の情報ブロックとして職業に関する情報ブロックに関する情報ブロックを選択し(ステップ170)、「職業が学生である」という検索条件を用いて、全てのレコードの中から職業が学生であるレコードを含む結果集合を作成する(ステップ172)。

【0045】

次に、「性別」に関する情報ブロックを第2の情報ブロックとして選択し、「年齢」に関する情報ブロックを第3の情報ブロックとして選択し(ステップ174)、結果集合の先頭から順次レコードへのポインタを取り出す(ステップ17

6)。

取り出されたレコードへのポインタを用いて、第2の情報ブロックの値管理テーブルへのポインタ配列を参照して取り出されたレコードへのポインタに対応した性別を獲得し、また、第3の情報ブロックの値管理テーブルへのポインタ配列を参照して取り出されたレコードへのポインタに対応した年齢を取り出す（ステップ178）。取り出された性別の男女別の総数を1ずつ増加し、取り出された年齢の合計を男女別に計算する（ステップ180）。

【0046】

結果集合の全てのレコードへのポインタが処理されたか否かを検査し（ステップ182）、全てのレコードへのポインタが処理された場合に、男女別に年齢の合計を総数で除算することにより学生の男女別の平均年齢を計算する（ステップ184）。

次に、図20に示された動作フローチャートを参照して、本発明の第7の実施例によるクロス集計方法について説明する。本実施例では、図1の表形式データに関してレコード全体を対象として、性別／職業別の存在数を求める場合を考える。

【0047】

集計に利用される性別及び職業の2個の項目について、夫々の値管理テーブルと、項目値番号指定情報配列、即ち、値管理テーブルへのポインタ配列とにより表された2個の個別項目情報、即ち、第1及び第2の情報ブロックを記憶装置に保持する（ステップ190）。記憶装置は、例えば、メモリ、仮想記憶、メモリマップドファイル等の形で実現される。

【0048】

性別に関する第1の情報ブロックについて、図6に示される如く性別の項目値番号の総数は2であり、また、職業に関する第2の情報ブロックについて、図8に示される如く、職業の項目値番号の総数は4であることが分かる。そこで、集計データを格納する空間として、 2×4 （2行4列）の2次元配列を初期化する（ステップ192）。

【0049】

第1及び第2の情報ブロックについて、夫々の値管理テーブルへのポインタ配列の先頭から順に夫々の項目値番号q1及びq2を取り出し、これにより、2次元配列の1個の要素P(q1, q2)を特定し(ステップ194)、特定された要素P(q1, q2)の値を1ずつ増加させる(ステップ196)。

値管理テーブルへのポインタ配列から全ての項目値番号(即ち、レコード総数個の項目値番号)が取り出されたか否かを検査し(ステップ198)、未だ項目値番号が残されているときにはステップ194に戻り、さもなければ、集計処理を終了する。このとき、2次元のクロス表が完成している。

【0050】

上記本発明の第7の実施例では、図1の表形式データのレコード全体を対象として集計を行っているが、レコードの部分集合を対象として同様の集計、例えば、年齢16歳の人の性別／職業別の存在数を集計することが可能である。そのため、最初に、年齢16歳を検索条件として単一項目検索を行い、年齢16歳に該当するレコードの識別情報を結果集合として取得、保持する。次に、既に、複数項目AND検索に関して説明しているように、この結果集合に含まれるレコード、即ち、レコードの部分集合を対象として、結果集合に含まれるレコードの識別情報を先頭から順次取り出し、レコードの識別情報に対応した項目値番号を値管理テーブルへのポインタ配列から取り出し、得られた項目値番号を2次元配列の行と列の座標とする要素の値のカウントを1ずつ加算すればよい。

【0051】

次に、項目の項目値が幾つかのカテゴリに分類されている状況において、項目値の分類毎に存在数をカウントする本発明の第8の実施例によるクロス集計方法について説明する。例えば、図8に示された「職業」に関する情報ブロックを参照するに、「職業」には、学生、プログラマ、教員、及び、その他の4種類の項目値が登録されていることが分かる。この項目値に基づく分類を、新たな基準、例えば、収入の有無に従って、所得者、無所得者及び不明の3タイプに再分類した場合を想定する。本例では、この状況において、「収入の有無」という新しいカテゴリを作成し、性別／収入の有無による存在数のクロス集計を作成する。

【0052】

図21に示された「職業」に関する情報ブロックは、特に、「収入の有無」に基づいて各項目値番号毎に分類番号が付与された値管理テーブルを含む。本例では、学生の分類番号は1（無所得）、プログラマ及び教員の分類番号は0（有所得）、その他の分類番号は2（不明）である。

本発明の第8の実施例のクロス集計方法は、原則として上記本発明の第7の実施例によるクロス集計方法と同様に動作するが、集計データを格納する2次元配列の要素を指定する座標として、性別に関する第1の情報ブロックの項目値番号と、職業に関する第2の情報ブロックの分類番号とを使用する点が相違する。

【0053】

2次元配列の要素の座標として項目値番号又は分類番号の何れでも使用できるようにするため、本発明の第8の実施例では、第1及び第2の情報ブロックについて、各値管理テーブルへのポイント配列に格納された夫々の項目値番号を順次取り出し、値管理テーブルへのポイント配列から取り出された項目値番号自体又はその項目値番号に対応して値管理テーブルに格納された分類番号の何れかで2次元配列の要素Pの座標を特定する。

【0054】

上記の本発明の第7及び第8の実施例によるクロス集計は、特に、存在数を求める形の集計であるが、本発明は、例えば、本発明の第9の実施例のクロス集計方法のように性別／職業別に平均年齢を求めるタイプのクロス集計にも拡張され得ることに注意する必要がある。

上記本発明の第7の実施例の如く、2次元配列の要素を1ずつインクリメントすることにより実現される存在数を求めるタイプの集計方法は、平均年齢を求めるような存在数の総和以外の演算を要するタイプのクロス集計に拡張するために図22の動作フローチャートに示されるように変形される。

【0055】

即ち、本発明の第9の実施例によれば、集計のために2個の2次元配列が用いられ、第1の2次元配列は上記第7の実施例と同様に性別／職業別の存在数をカウントし、第2の2次元配列は性別／職業別の年齢の総和を計算する。詳細に説明すると、

性別、職業及び年齢の3個の項目について、第1及び第2の情報ブロックを記憶装置に保持し（ステップ200）、

性別及び職業の項目値番号の総数2及び4に対応して、集計データを格納する2個の 2×4 （2行4列）の2次元配列を初期化し（ステップ202）、

第1及び第2の情報ブロックの値管理テーブルへのポインタ配列の先頭から順に項目値番号 q_1 、 q_2 及び q_3 を取り出し、2次元配列の要素の座標（ q_1 , q_2 ）を特定し（ステップ204）、

第1の2次元配列の要素 P_1 （ q_1 , q_2 ）の値を1ずつ増加し（ステップ206）、

「年齢」に関する情報ブロックについて、項目値番号 q_3 に対応した項目値（即ち、年齢）を取得し（ステップ208）、

第2の2次元配列の要素 P_2 （ q_1 , q_2 ）の値に取得された年齢を加算し（ステップ210）、

対象となる全てのレコードが処理されたか否かを検査し（ステップ212）、処理されていない場合にステップ204に戻り、

処理されている場合に、2次元配列 P_1 と2次元配列 P_2 の各要素の間で演算、即ち、

$$P_2(q_1, q_2) / P_1(q_1, q_2)$$

の計算を行う（ステップ206）。これにより、性別／職業別の平均年齢が得られ、平均のクロス集計表が作成される。

【0056】

図23の（A）は、上記本発明の第7の実施例で得られるクロス集計表の概念図である。このように、上記第7の実施例では、全ての性別／職業の組合せに関する存在数が集計される。しかし、図23の（B）に示される如く、性別／職業別の中で、例えば、特に、性別が女性で、職業が学生である人の存在数のみを知りたい場合がある。本発明によれば、この場合の存在数は、「女性」AND「学生」という複数項目AND検索の結果集合のサイズを求めることにより獲得される。

【0057】

同様に、上記本発明の第9の実施例では、全ての性別／職業の組合せに関して、平均年齢のクロス集計表を求めているが、特に、性別が女性で、職業が学生である人の平均年齢だけを求めることが可能である。その場合、「女性」AND「学生」という複数項目AND検索の結果集合のサイズから存在数を求めると共に、その結果集合に含まれるレコードの識別情報によって指定されるレコードに属する年齢を加算して年齢の総和を求め、(年齢の総和)／(存在数)を計算することにより平均年齢のクロス集計表の特定のセルの値を求めることができる。

【0058】

図24はマルチアンサ形式の項目を表す図であり、図25は本発明の第10の実施例によるマルチアンサ項目対応型の情報ブロックの説明図である。マルチアンサとは、例えば、「今、机の上にある文房具は何ですか?」という質問に対する回答を得た場合に「鉛筆、消しゴム」又は「紙、鉛筆」のように同じ人から複数の回答が得られる状況を意味する。即ち、マルチアンサの場合、一つのレコードが一つの項目に関して複数の項目値を指定することが生じ得る。図24は上記の質問に対し、100万人から得た回答をそのまま一覧表にしたものである。

【0059】

このようなデータを処理するため、本発明の第10の実施例によれば、図25に示される如く、情報ブロックの値管理テーブルへのポインタ配列は、前述の項目値番号自体の配列ではなく、項目値番号毎に1ビットずつが割り当てられ、当該レコードがその項目値番号を指定しているか否かをビットのオン／オフによって指定する2進値の配列により構成する。これにより、一つのレコードについて、一つの項目の複数の項目値を指定することが可能になる。本発明によれば、情報ブロックの一部分の構成を変形するだけでマルチアンサに容易に対応することができる点が有利である。實際上、このような変形を加えた情報ブロックによって上記の本発明の種々の実施例で採用された情報ブロックを置き換えることができる。

【0060】

図26は、本発明の第11の実施例に従って、集計処理の際に生じるブランク、エラー値等の特殊値の取扱方法を説明する図である。同図にはブランクをカテ

ゴリーの1つとみなしてクロス集計した例が示されている。実際のデータを取り扱う場合には、ブランクや、 $\log(-1)$ のような数学的なエラーが出現することがある。本発明によれば、このような特殊値であっても、ブランク、エラー等を値管理テーブルに項目値として登録し、登録された特殊値をそのまま検索や集計のカテゴリーとして使用できる点が有利である。

【0061】

次に、図27に示された本発明の第12の実施例による複数項目検索方法の動作フローチャートを参照して遅延評価について説明する。本実施例は、上記の本発明の第2の実施例と同様に、図1に示した表形式データに関して「年齢」が16歳又は19歳であるという第1の検索条件と、「職業」が学生であるという第2の検索条件を満たすレコードの集合を得る場合を考える。

【0062】

上記本発明の第2の実施例では、第2の検索条件に関して、全てのレコードの分類番号を予め設定しているが（図12のステップ124）、本発明の第12の実施例の場合、分類番号の設定は、第1の検索条件による検索の結果集合に基づいて実際にアクセスされた項目値番号に対応した分類番号だけに対して行われる。

【0063】

既に説明したように、第1の項目である「年齢」に関する情報ブロックである第1の特定情報ブロックについて、「年齢」が16歳又は19歳であるレコードの結果集合は本発明の第1の実施例に従って得られる（ステップ220）。

次に、図8に示されている第2の項目である「職業」に関する情報ブロックである第2の特定情報ブロックとして選択し（ステップ222）、第2の特定情報ブロックの値管理テーブルの全ての分類番号の値を例えば、-1に初期化する（ステップ224）。

【0064】

次に、第1の検索条件に関する結果集合から、レコードへのポインタを表すレコード番号を順次取り出す（ステップ226）。本例の場合、図14に示されるように、例えば、レコード番号999999が取り出される。

次に、第2の特定情報ブロックに関して、値管理テーブルへのポインタ配列から上記第1の検索条件について得られたレコード番号に対応した項目値番号を取り出す（ステップ228）。本例の場合、図14に示されるように、例えば、レコード番号999999に対応した項目値番号0が取り出される。

【0065】

次に、第2の特定情報ブロックについて取り出された項目値番号に対応する分類番号の値が-1であるか否かを検査する（ステップ230）。

分類番号が-1の場合には、当該項目値番号に対して未だ分類番号が設定されていないので、この項目値番号に対応した項目値が上記第2の検索条件に合致するかどうかを判定し（ステップ232）、合致する場合には、分類番号として1を設定し（ステップ234）、合致しない場合には、分類番号として0を設定する（ステップ236）。

【0066】

分離番号の値が-1ではない場合、上記取り出された項目値番号に対応する分類番号の値が1に設定されているかどうかを判定する（ステップ238）。分類番号の値が1に設定されている場合、値管理テーブルへのポインタ配列内で、分類番号が1にセットされている当該項目値番号を指すポインタが格納されている場所に対応したレコードへのポインタ、例えば、レコード番号を最終的な結果集合に追加する（ステップ240）。本例では、図14に示されているように、例えば、レコード番号999999が最終的な結果集合に追加される。分類番号が0である場合には、最終的な結果集合は更新されない。

【0067】

また、本発明の情報ブロックの一実施例によれば、電話番号のように国番号+市外局番+局番+番号のような構造を持つデータを複数の情報ブロックに分割して登録することにより、国番号、市外局番等の一部のデータに関する検索・集計が容易に行える利点を得られる。

また、既に説明した本発明の一実施例による分類番号を利用することにより、例えば、年齢に関して、10～19歳を10代とし、20～29歳を20代とし、以下同様に取りまとめて新たなカテゴリー区分を生成することができ、かつ、

生成された新しいカテゴリーに関して、前述と同様の検索・集計方法を適用できる。

【0068】

図28は本発明の一実施例による表形式データの検索・集計装置が実現されるシステムの構成図である。同図に示す如く、表形式データの検索・集計装置は、一般的なコンピュータシステム、例えば、CPU100と、ROM110と、RAM120と、ハードディスク130と、ディスプレイ140のような出力装置と、KB/マウス150等の入力装置とがバス160を介して接続されているパーソナルコンピュータ上に実装することが可能である。表形式データの検索・集計装置は、パーソナルコンピュータ上で実行されるソフトウェアの形態で提供され、情報ブロックは、RAM120及びハードディスク130のような記憶装置上に展開される。

【0069】

次に、図29に示されたフローチャートを参照して、一例として図2の(B)に示された表形式データに対し、図5に示された形式の情報ブロックを構築する方法を説明する。

ステップ300：データの準備

まず、図2の(B)に示す形態のデータを用意する。次に、これを項目毎に分解する。図2の(B)では、「性別」、「年齢」及び「職業」の項目に分解することができる。

【0070】

ステップ311：「性別」項目用の情報ブロックの生成

1つの情報ブロックを生成し、これを、例えば、「性別」項目の情報ブロックとする。

ステップ312：値管理テーブルの作成

次に、値管理テーブルを初期化し、「性別」項目データを先頭から末尾まで走査することで、どのような項目が何件あるかを数え、格納する。本例の場合、ここまでの段階で、上記値管理テーブルの項目値11には、「女性」及び「男性」がセットされ、上記値管理テーブルの存在数14には、367436、6325

64がセットされることになる。

【0071】

次に、項目値11（「女性」及び「男性」）を所望の基準に従ってソートする。もちろん、そのソートの際には、存在数14も項目値11と共に並び替えねばならない。

さらに、値管理テーブルの開始位置13をセットする。この開始位置13は、値管理テーブル内の最初の存在数14から、該開始位置13の値管理テーブル内の1つ前までの存在数14の累計として求められる。もちろん、先頭の開始位置13の値は、0である。

【0072】

次に、開始位置13の内容を、分類番号12にコピーする。分類番号12は、後程、レコードへのポインタ配列を作成する際の作業領域として使用する。

ステップ313：値管理テーブルへのポインタ配列の作成

次に、値管理テーブルへのポインタ配列20のための領域を確保する。（領域のサイズは、上記存在数14の総和である。）

さらに、「性別」項目データを先頭から末尾まで、項目値を1つずつ取り出し、各項目値が、値管理テーブルの何番目の項目値と一致するかを調べ、n番目の項目値であった場合、n-1を値管理テーブルへのポインタとして上記値管理テーブルへのポインタ配列に格納する。

【0073】

ステップ314：レコードへのポインタ配列の作成

次に、レコードへのポインタ配列30のための領域を確保する。本例では、領域のサイズは、上記存在数14の総和である。値管理テーブルへのポインタ配列20の先頭から末尾まで、1つずつ値管理テーブルへのポインタを取り出す。値管理テーブルへのポインタ配列20のJ番目の値を取り出し、その値がKであり、値管理テーブルの(K+1)番目のレコードの分類番号12の値を取り出し、その値がLであったとすると、レコードへのポインタ配列30の(L+1)番目の要素にJ-1を格納し、値管理テーブルの(K+1)番目のレコードの分類番号12の値を1だけ増加させる。

【0074】

以上の操作で、「性別」項目の情報ブロックの作成が完了した（ステップ310）。「年齢」項目、「職業」項目も同様に情報ブロックを作成することができ（ステップ320及びステップ330）、表形式データの全体に対する情報ブロックが得られる。

図30乃至35は、図1に示された表形式データの「職業」に関する情報ブロックを作成する手順の説明図である。

【0075】

図30には、属性値の種類が予め分かっているカテゴリー定義済みの場合の新規データ投入を示す図である。ここでは、既知のカテゴリー定義に従って、値管理テーブルが作成される。開始位置、存在数は不明であるため、0に初期化する。また、値管理テーブルへのポインタ配列、レコードへのポインタ配列の為の領域を確保し、初期化しておく。

【0076】

図31には、値管理テーブルへのポインタ配列と、値管理テーブルの存在数を完成させるパスが示されている。投入したいデータを先頭からひとつずつ取りだし、それが値管理テーブルの何番目の値であるか（即ち、項目値番号）を調べ、値管理テーブルへのポインタ配列に格納し、値管理テーブルの存在数を+1ずつ更新する。図31の例は、投入データの上から2番目までの処理が終わった状態を示している。

【0077】

図32には、値管理テーブルを完成させる第2のパスが示されている。存在数の累計は開始位置に対応することを利用して、開始位置が求められる。さらに、分類番号には開始位置の値をコピーする。同図では分類番号の設定が終了している。

図33乃至35には、データ投入の第3のパスが示されている。このパスは、値管理テーブルへのポインタ配列の先頭から1つずつ値を取りだし、その値が参照する値管理テーブル内の分類番号の指示するレコードへのポインタ配列に、値管理テーブルへのポインタ配列のオフセット、即ち、レコード番号を格納する。

図33、34及び35には、夫々、「職業」に関する情報ブロックの値管理テーブルへのポインタ配列の先頭、2番目及び最後のデータについての処理が示されている。

【0078】

尚、上記の説明では、分類番号のフィールドが作業領域として使用されているが値管理テーブルの行数、即ち、項目値番号の総数と同数以上の要素を持つ整数配列であればいかなる配列でも作業領域として使用することができる。

一方、予めカテゴリー定義がなされていない場合の新規投入は、投入データを走査して値管理テーブルに登録すべき値の一覧を獲得した後、上記のカテゴリー定義がある場合の新規投入の処理を行うことにより実現される。

【0079】

次に、図35に示されるような新規投入が完了した後の「職業」に関する情報ブロックに対し、「学生」という別のデータを追加する場合を考える。図36はかかるデータ追加を説明する図である。

この場合、最初に、値管理テーブルへのポインタ配列の末尾に「学生」を表す項目値番号0を追加し、値管理テーブルの中の学生の存在数を1増やす。次に、値管理テーブルへのポインタ値(=1000000)、即ち、レコード番号を格納するための領域をレコードへのポインタ配列内に確保する必要がある。そのため、レコードへのポインタ配列の中の「学生」に対応した領域の末尾の値(本例では、999999)を取り出し、そこに、拡張先のアドレスである1000000を格納する。但し、拡張先のアドレスであることが識別できるように、例えば、符号を反転させて、-1000000として格納する。続いて、拡張先に先に取り出された末尾の値である999999を格納し、最後に、新たに追加されたデータに対応したポインタ値1000000を格納する。

【0080】

このような追加方法を採用すると、データを追加する際に大量のデータを移動させる必要が回避される。また、多数のデータ追加が行われた場合に、ポインタの数が増加することに起因したアクセスの効率低下を抑えるために、適当なタイミングで新規投入の第3のパスと同じ処理を繰り返しても良い。

図37には、本発明の他の実施例による情報ブロックの構造が示されている。同図に示された構造を採用した場合、ポイントの数の増加は回避され容易にデータ変更を行うことが可能である。

【0081】

【実施例】

次に、100万件のデータの検索・集計のテストについて説明する。テストに使用したプラットフォームは、Pentium Pro（登録商標）200MHzと、128MBメモリを搭載した一般的なパーソナルコンピュータである。図38はテストに使用したデータの一覧を示す図表である。000000～999999までの100万個の数字を、10000の単位、100の単位、1の単位の3項目に分解した表形式データである。各項目に00～99までの項目値が1万回ずつ出現する。

【0082】

図39は、100万件の検索／集計に要する時間を使用した結果集合のタイプ毎に測定した測定結果の一覧である。結果集合のタイプは、上記のビットフラグ形式とポイント配列形式の2種類である。測定結果の時間はミリ秒（1/1000秒）単位である。

上記テストで行った検索は、×1万、×100、×1の3項目をAND条件で接続した複数項目AND検索である。検索は、×1万、×100、×1の項目順にカスケードした。検索の中間及び最終の結果集合は、上記の通りビットフラグ又はポイント配列のいずれかの形式である。測定時間は5回の測定の平均である。

【0083】

また、上記テストの集計は、検索のテストにより得られた結果集合に対して、×100、×1の各値（00～99）が出現した回数の計数である。このクロス集計の表のサイズは、100×100の1万セルである。時間は全てミリ秒単位であり、5回の測定の平均を測定時間とした。

また、表形式データの検索・集計システムの構成は、上記の実施例で説明された例に限定されることなく、検索・集計システムの各々の構成要件をソフトウェア（プログラム）で構築し、ディスク装置等に格納しておき、必要に応じて検索

・集計システムのコンピュータにインストールして表形式データの検索・集計を行うことも可能である。さらに、構築されたプログラムをフロッピーディスクやCD-ROM等の可搬記録媒体に格納し、このようなシステムを用いる場面で汎用的に使用することも可能である。

【0084】

本発明は、上記の実施例に限定されることなく、特許請求の範囲内で種々変更・応用が可能である。

【0085】

【発明の効果】

上記の説明の通り、本発明によれば、表形式で表現される大量のデータを、アクセスに時間がかかる従来のデータ表を用いることなく処理することができるので、集計・検索のスピードを大幅に向上させることができる。

【図面の簡単な説明】

【図1】

表形式データの説明図である。

【図2】

(A) 及び (B) は、従来技術による表形式データの格納技術の説明図である。

【図3】

本発明の原理説明図である。

【図4】

本発明による情報ブロックの説明図である。

【図5】

本発明による情報ブロックの説明図である。

【図6】

本発明の実施例で使用される「性別」に関する情報ブロックの説明図である。

【図7】

本発明の実施例で使用される「年齢」に関する情報ブロックの説明図である。

【図8】

本発明の実施例で使用する「性別」に関する情報ブロックの説明図である。

【図 9】

本発明の第 1 の実施例による単一項目内の検索方法の動作フローチャートである。

【図 10】

本発明の第 1 の実施例による情報ブロックの説明図である。

【図 11】

本発明の第 1 の実施例による情報ブロックの説明図である。

【図 12】

本発明の第 2 の実施例による複数項目の AND 検索方法の動作フローチャートである。

【図 13】

本発明の第 2 の実施例による情報ブロックの説明図である。

【図 14】

本発明の第 2 の実施例による情報ブロックの説明図である。

【図 15】

本発明の第 3 の実施例によるビットフラグを用いた複数項目論理演算検索方法の説明図である。

【図 16】

本発明の第 3 の実施例によるビットフラグを用いた複数項目論理演算検索方法の説明図である。

【図 17】

本発明の第 5 の実施例による集計方法の動作フローチャートである。

【図 18】

本発明の第 6 の実施例の概念的な説明図である。

【図 19】

本発明の第 6 の実施例の動作フローチャートである。

【図 20】

本発明の第 7 の実施例によるクロス集計の動作フローチャートである。

【図21】

本発明の第8の実施例による情報ブロックの説明図である。

【図22】

本発明の第9の実施例によるクロス集計の動作フローチャートである。

【図23】

(A) 及び (B) はクロス集計表の概念説明図である。

【図24】

マルチアンサ形式の項目の説明図である。

【図25】

本発明の第10の実施例によるマルチアンサ項目対応型の情報ブロックの説明図である。

【図26】

本発明の第11の実施例による特殊値の取扱方法の説明図である。

【図27】

本発明の第12の実施例による複数項目検索方法の動作フローチャートである。

【図28】

本発明の一実施例による表形式データの検索・集計システムの構成図である。

【図29】

情報ブロックの構築方法の説明図である。

【図30】

データ投入の準備と初期化の説明図である。

【図31】

データ投入の第1のパスの説明図である。

【図32】

データ投入の第2のパスの説明図である。

【図33】

データ投入の第3のパスの説明図である。

【図34】

データ投入の第3のパスの説明図である。

【図35】

データ投入の第3のパスの説明図である。

【図36】

情報ブロックに対するデータ追加の説明図である。

【図37】

本発明の他の実施例による情報ブロックの構造を示す図である。

【図38】

検索・集計テストに使用した100万件のデータの説明図である。

【図39】

100万件のデータの検索・集計テストの測定結果を示す図である。

【符号の説明】

- 10 値管理テーブル
- 11 項目値
- 12 分類番号
- 13 開始位置
- 14 存在数
- 20 値管理テーブルへのポインタ配列
- 30 レコードへのポインタ配列

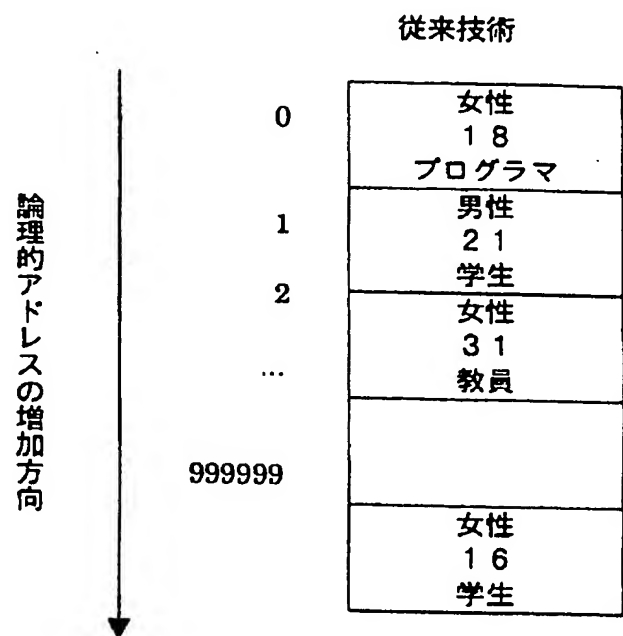
【書類名】 図面

【図1】

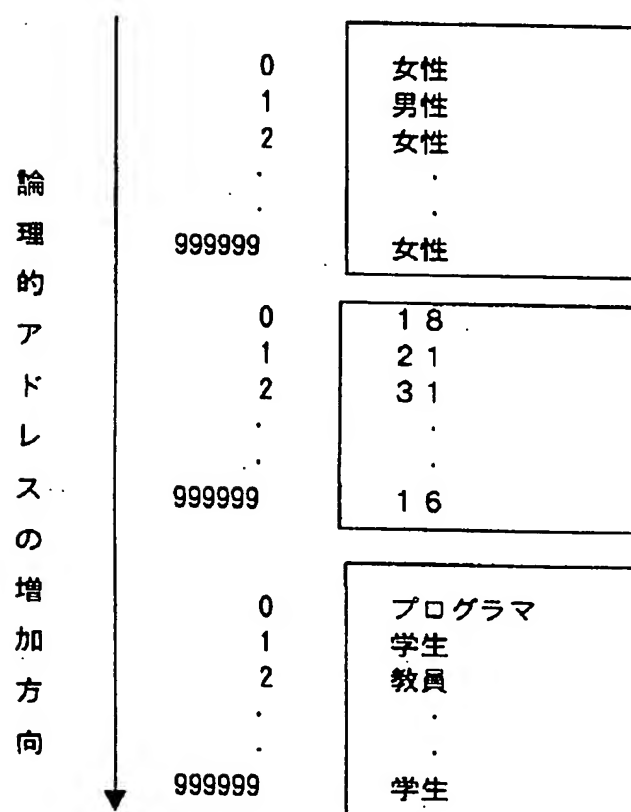
表形式データ

| レコード番号 | 性別 | 年齢 | 職業 |
|--------|----|----|-------|
| 0 | 女性 | 18 | プログラマ |
| 1 | 男性 | 21 | 学生 |
| 2 | 女性 | 31 | 教員 |
| | | | |
| 999999 | 女性 | 16 | 学生 |

【図2】



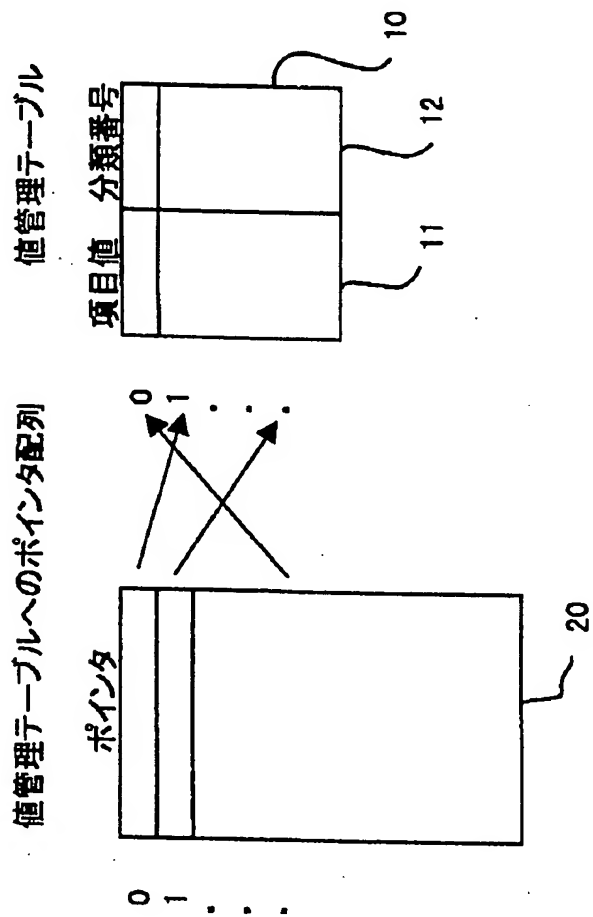
(A)



(B)

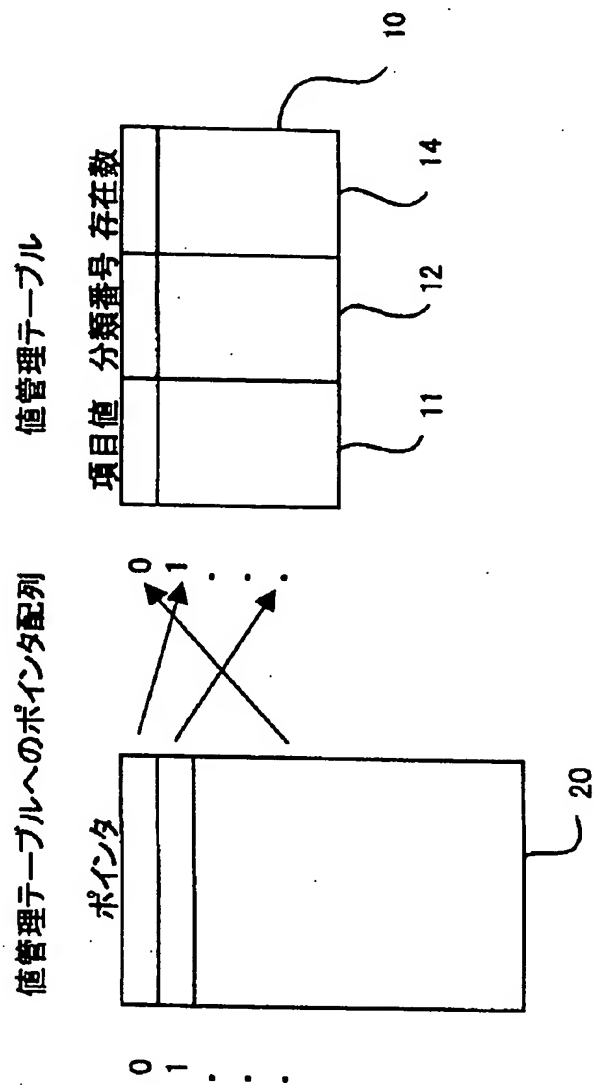
【図3】

本発明の原理説明図



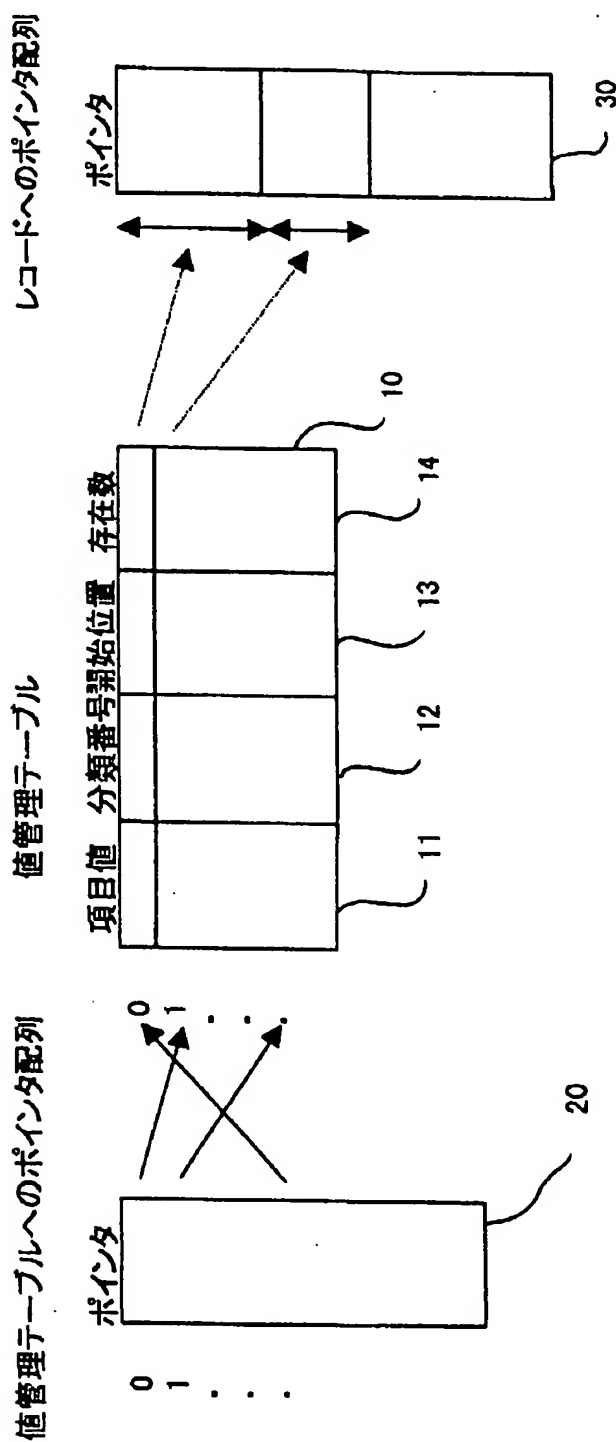
【図4】

本発明による情報ブロックの説明図



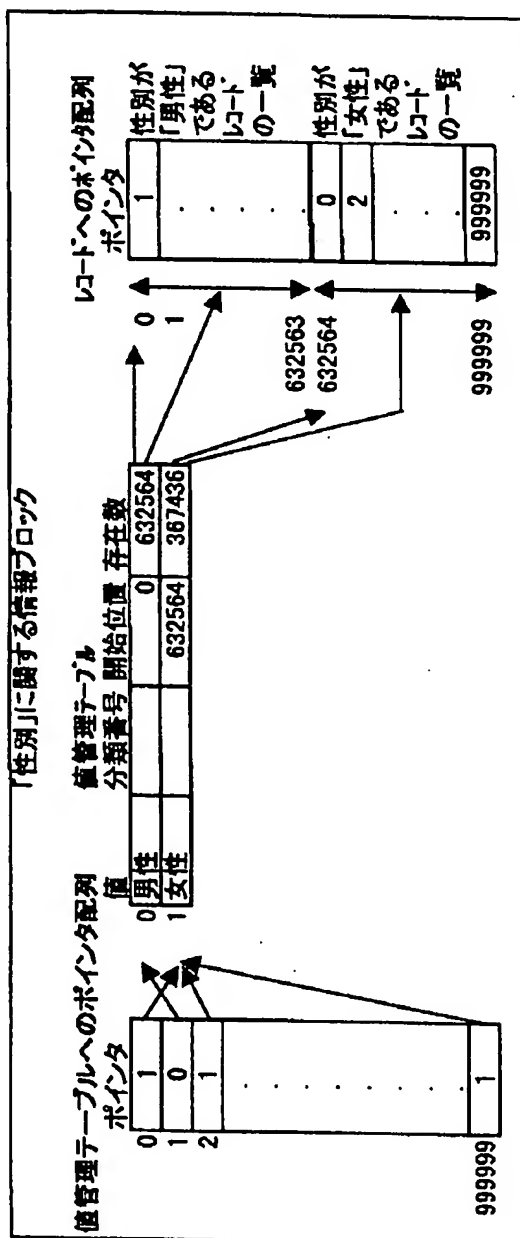
【図5】

本発明による情報ブロックの説明図



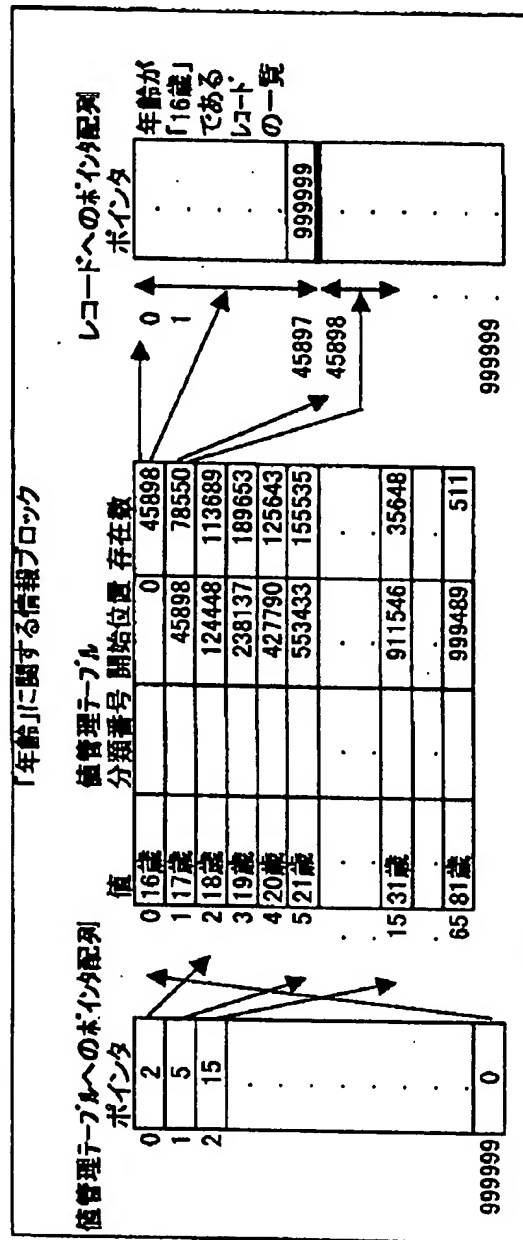
【図 6】

本発明の好ましい一実施例による情報ブロックの説明図



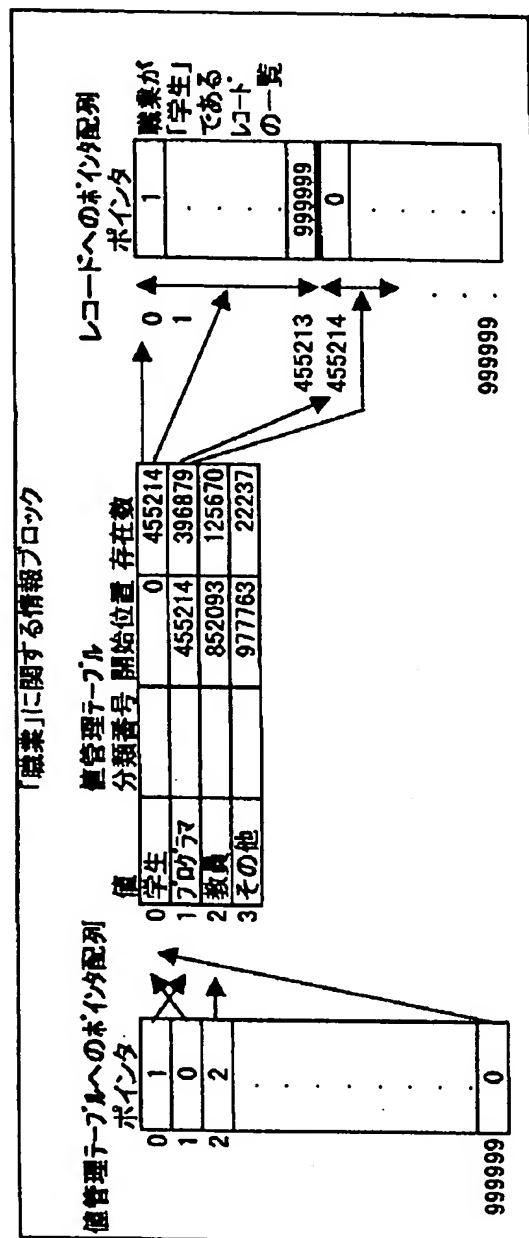
【図7】

本発明の好ましい一実施例による情報ブロックの説明図



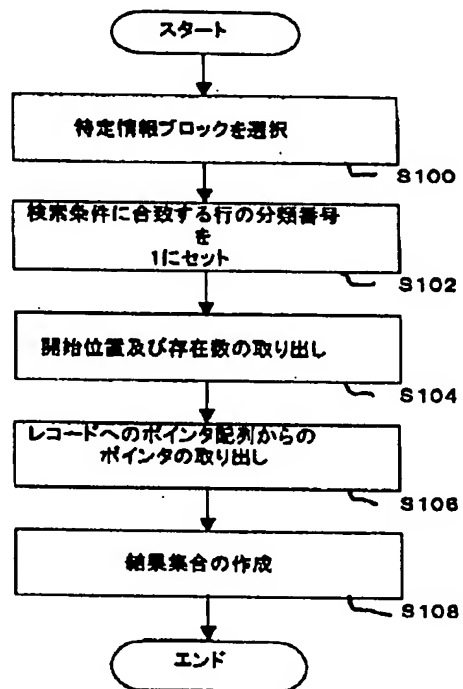
【図8】

本発明の好ましい一実施例による情報ブロックの説明図



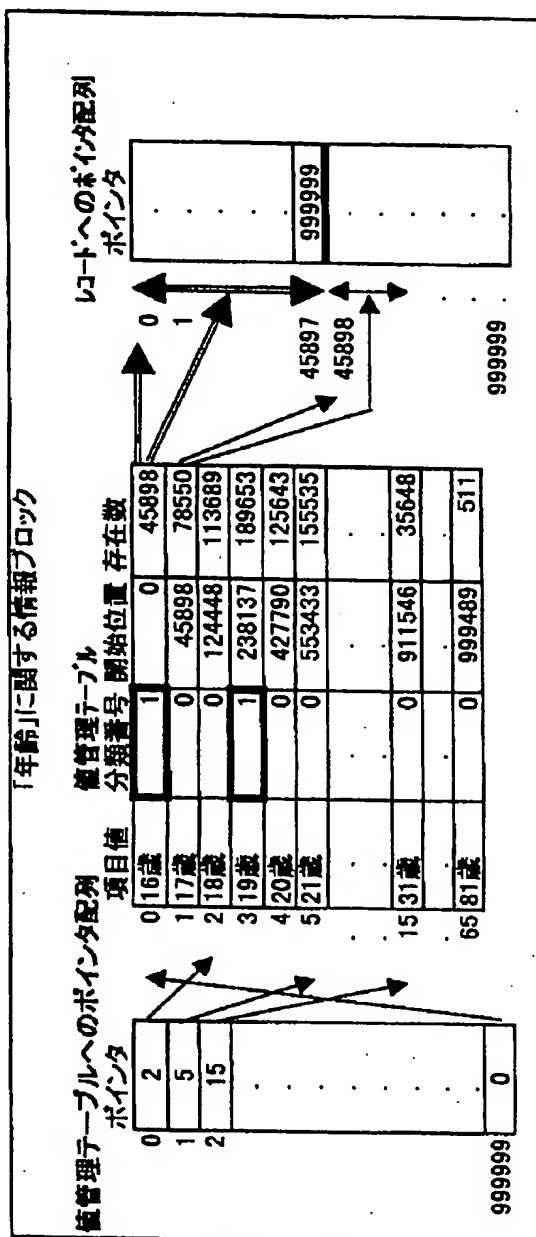
【図9】

本発明の第1の実施例による単一項目内の検索方法の動作フローチャート



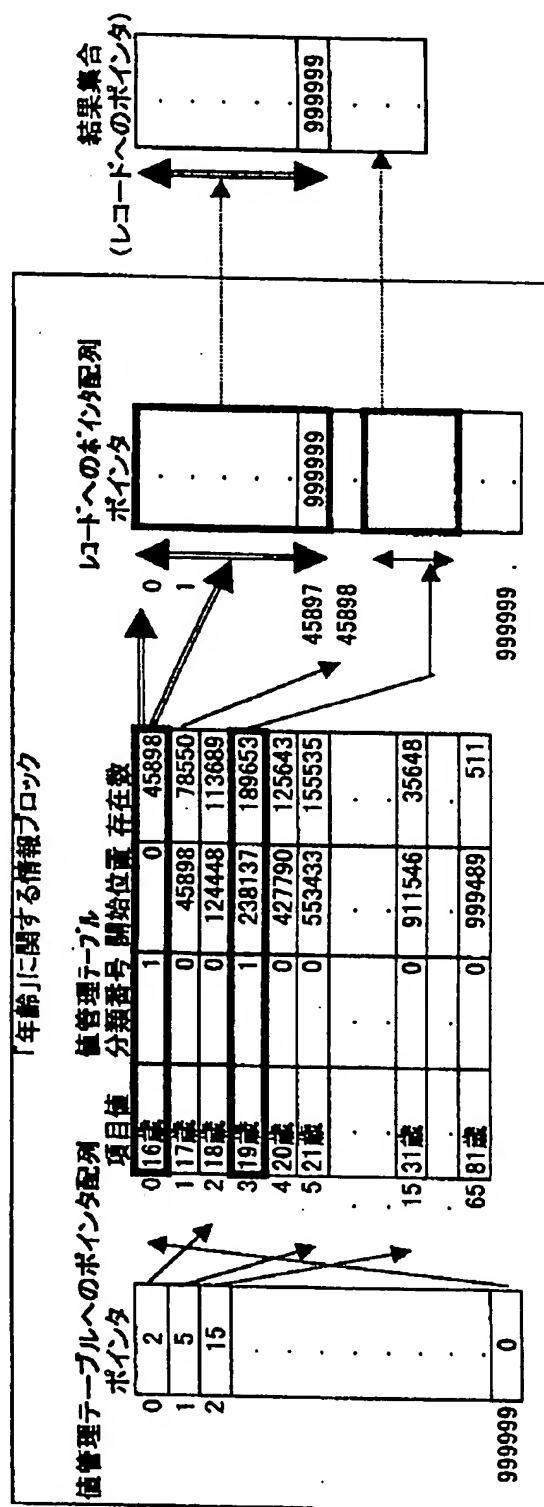
【図 10】

本発明の第1の実施例による単一項目検索の説明図



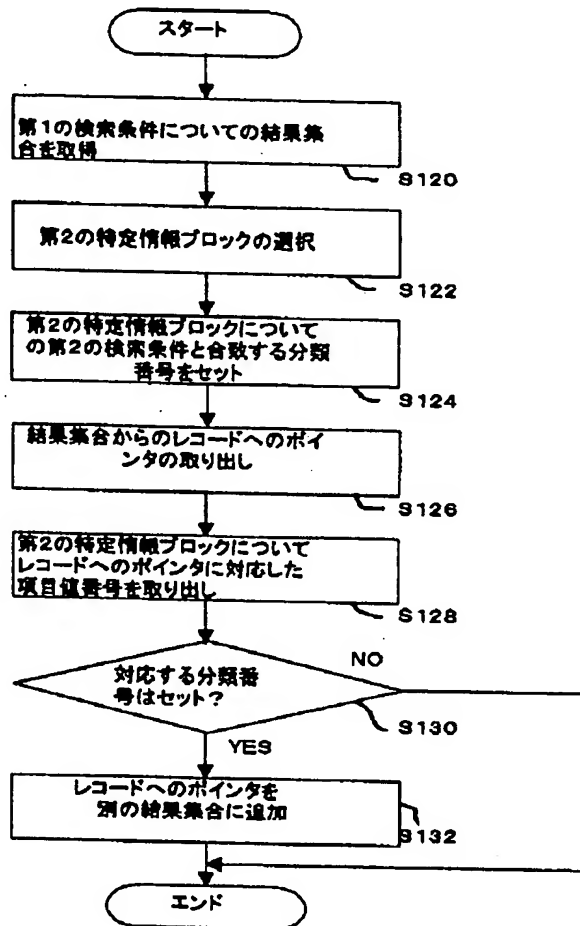
【図11】

本発明の第1の実施例による単一項目検索の説明図



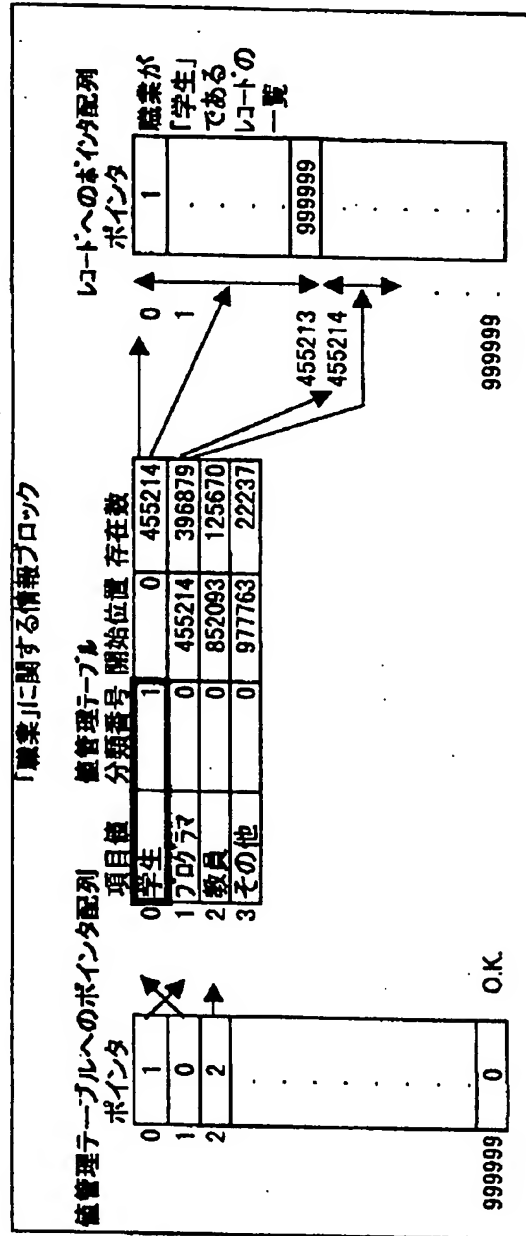
【図12】

本発明の第2の実施例による複数項目AND検索方法の動作フローチャート



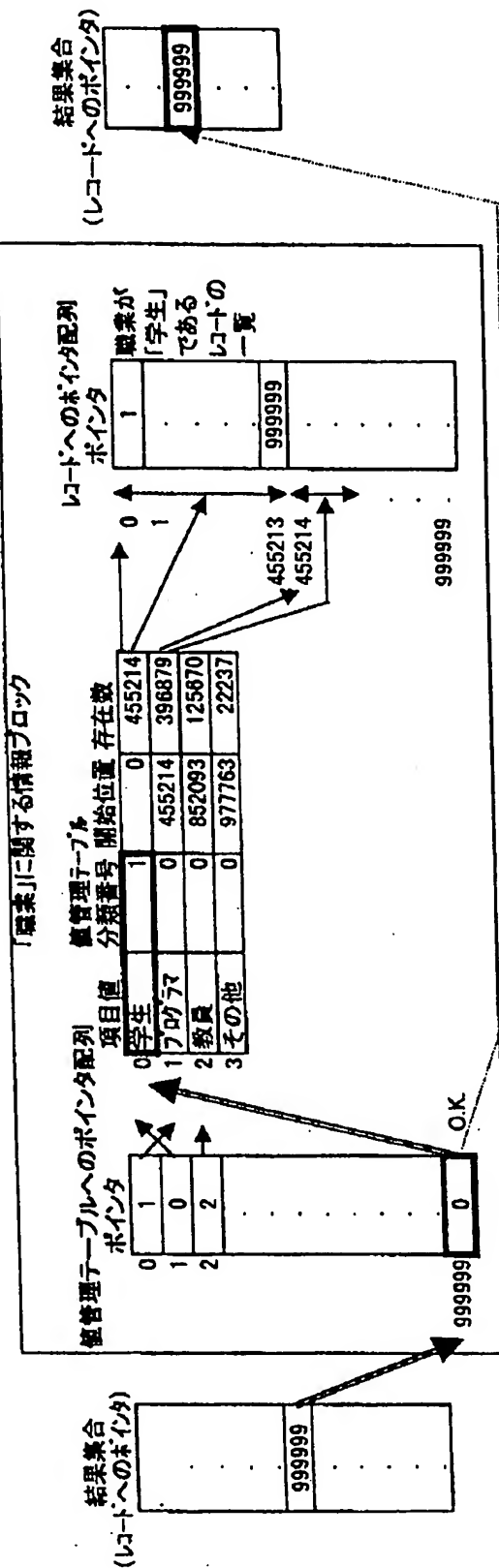
【図13】

本発明の第2の実施例による複数項目AND検索方法の説明図



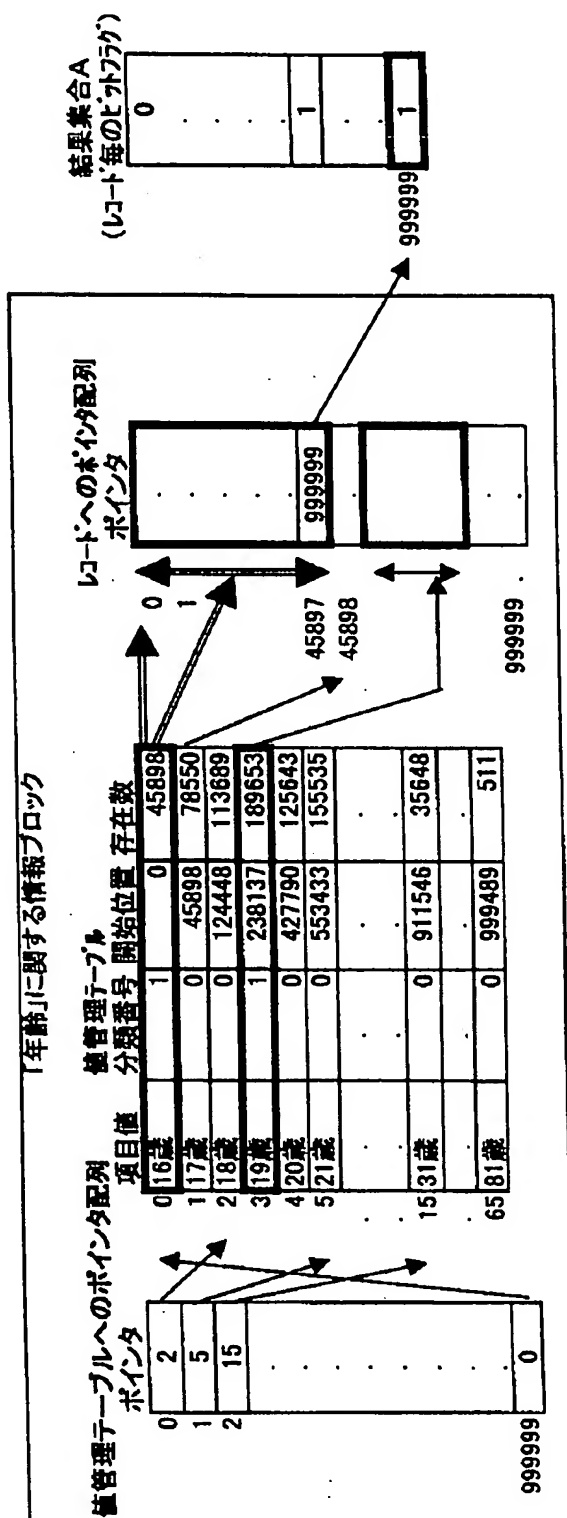
【図14】

本発明の第2の実施例による複数項目AND検索方法の説明図



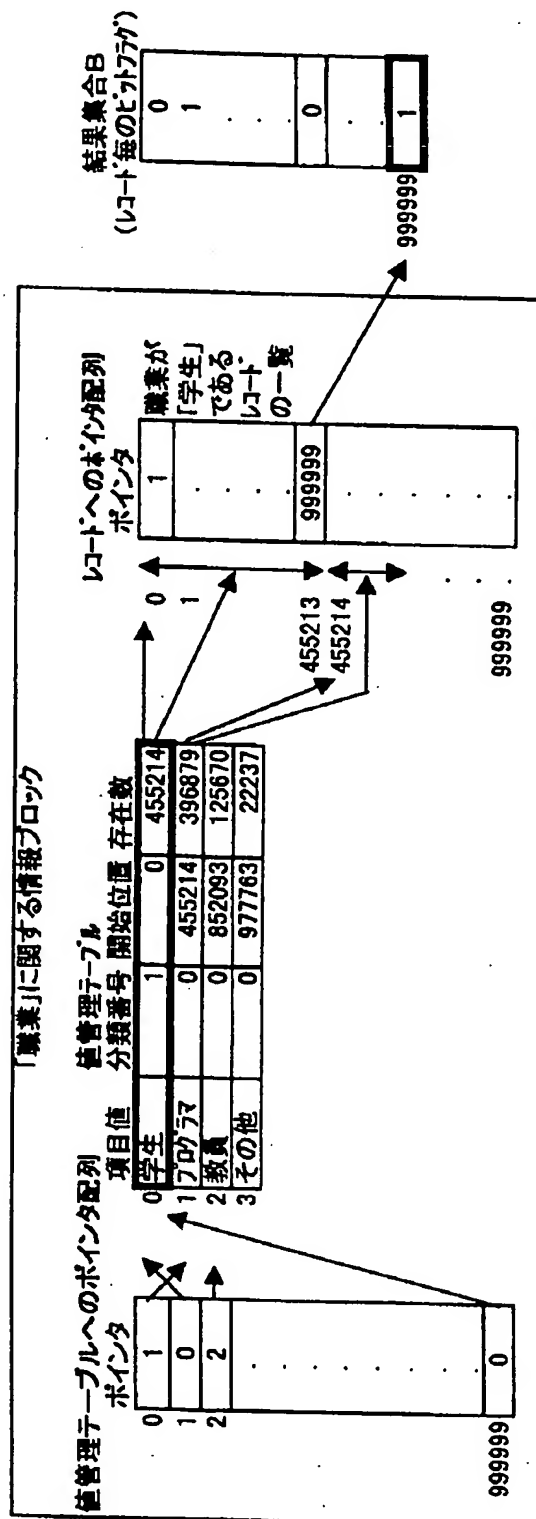
【图 15】

本発明の第3の実施例によるビットフラグAND検索の説明図



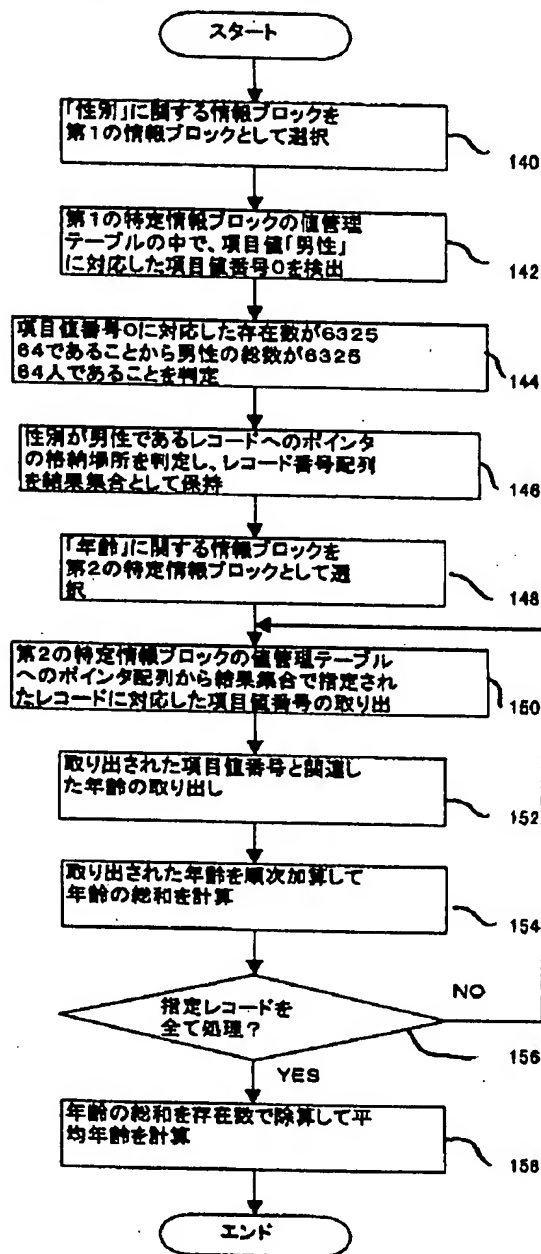
【図16】

本発明の第3の実施例によるビットフラグAND検索方法の説明図



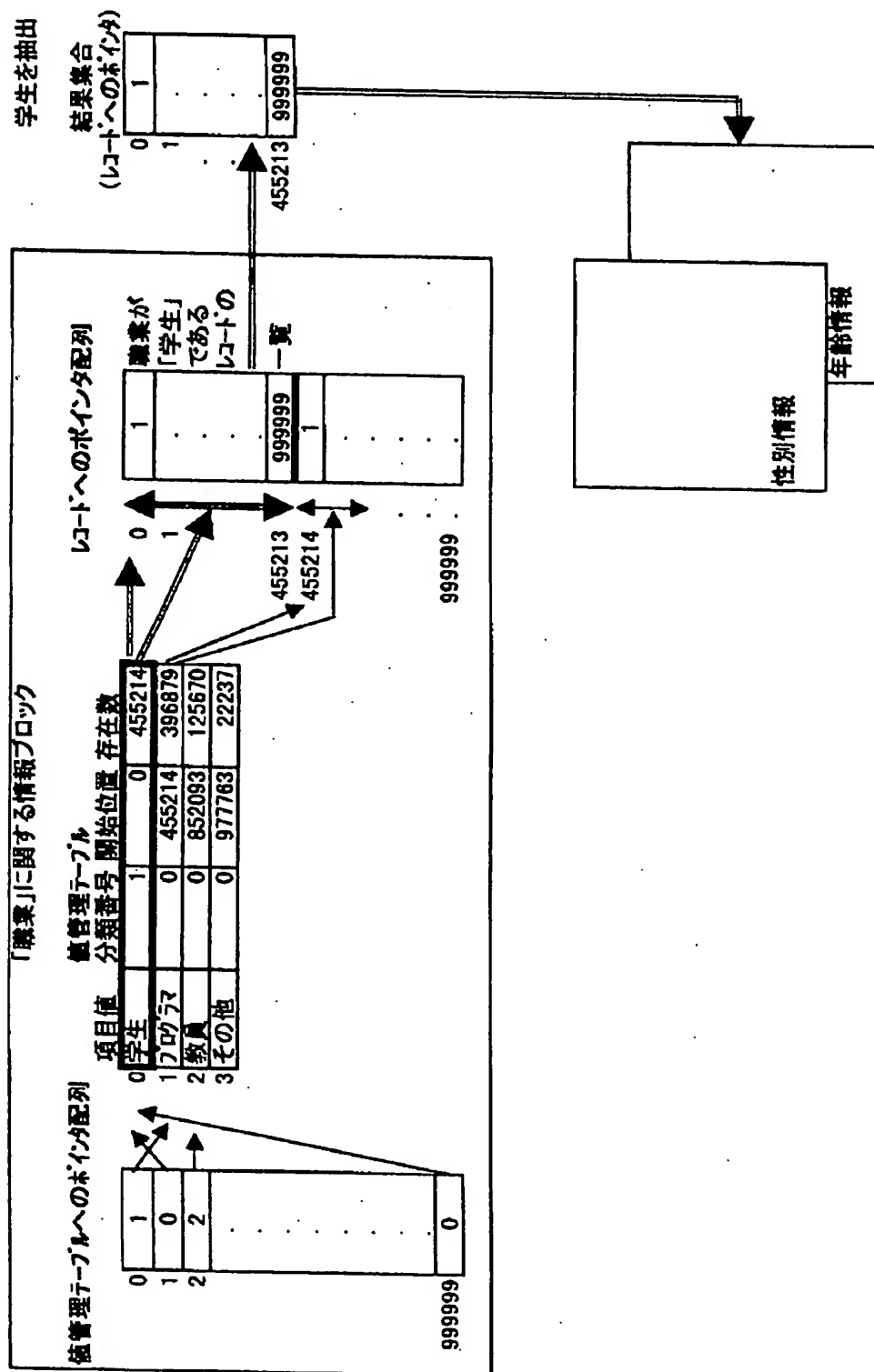
【図17】

本発明の第5の実施例の動作フローチャート



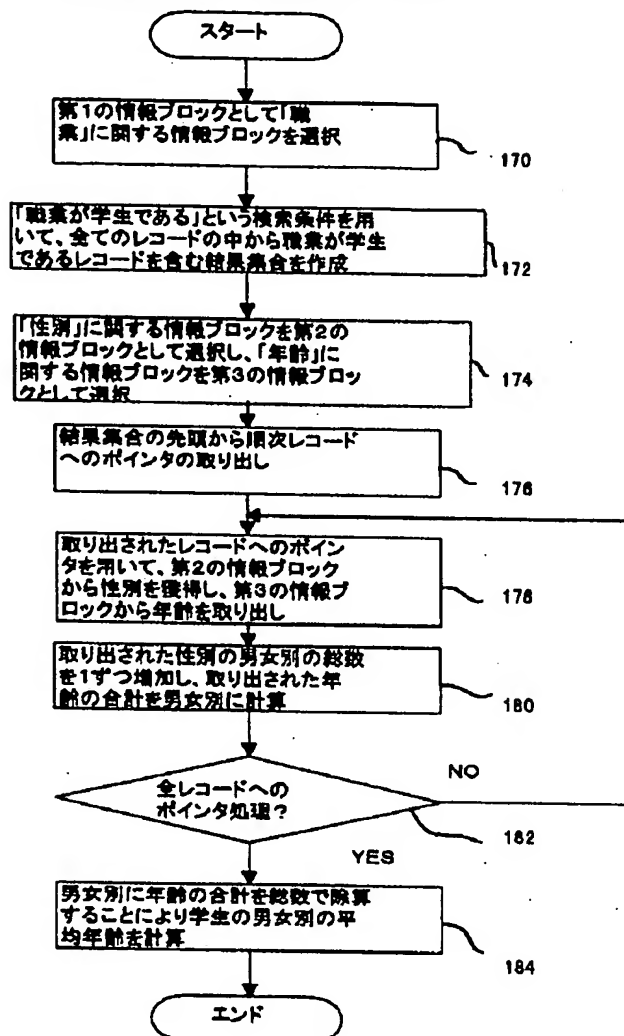
【図18】

本発明の第6の実施例の説明図



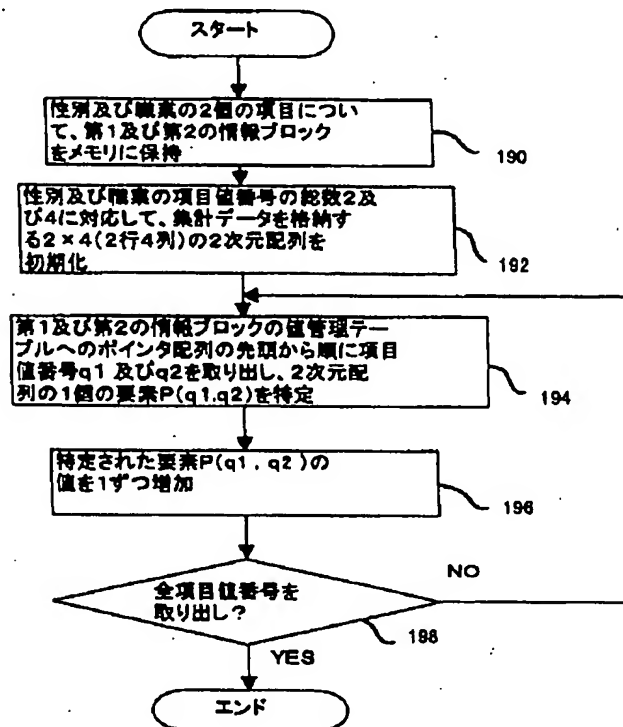
【図19】

本発明の第6の実施例の動作フローチャート



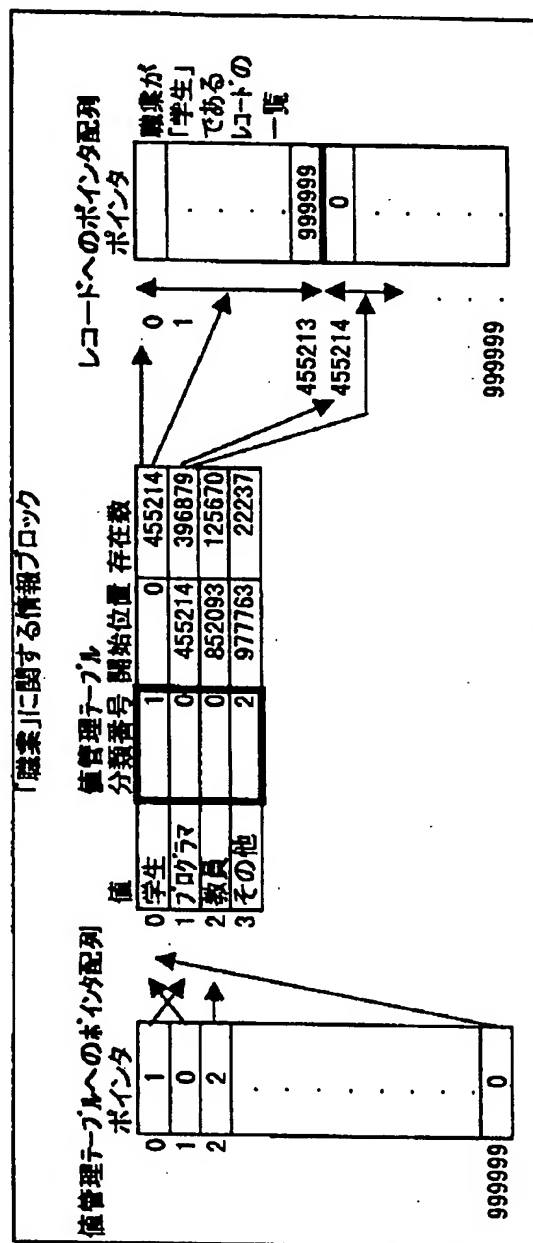
【図20】

本発明の第7の実施例によるクロス集計方法の動作フローチャート



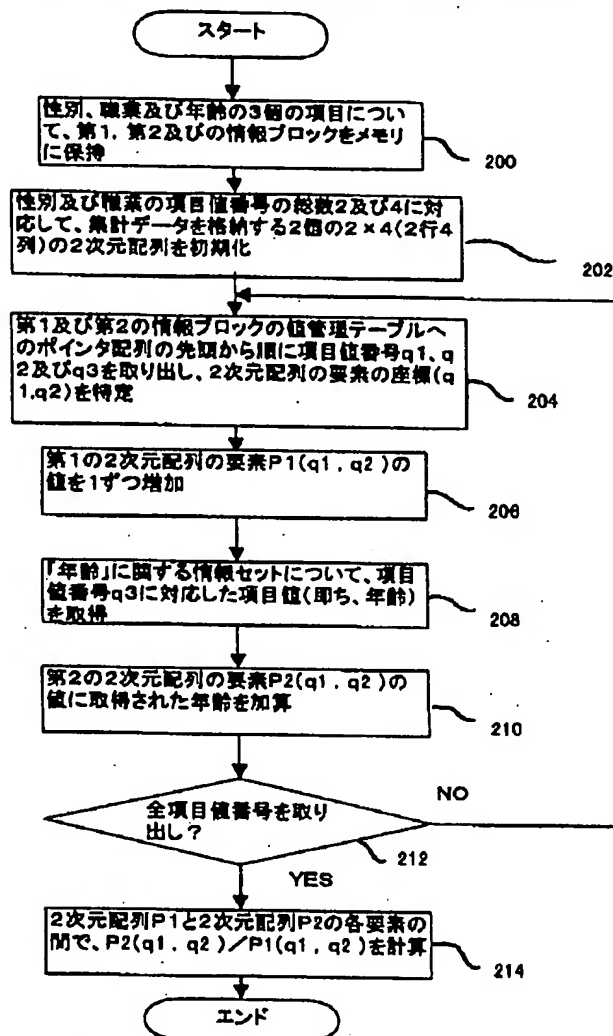
【図21】

本発明の第8の実施例による情報ブロックの説明図



【図22】

本発明の第9の実施例によるクロス集計方法の動作フローチャート



【図23】

クロス集計表の例の説明図

| | 学生 | プログラマ | 教員 | 無職 |
|----|----|-------|----|----|
| 男性 | ○ | ○ | ○ | ○ |
| 女性 | ○ | ○ | ○ | ○ |

(A)

| | 学生 | プログラマ | 教員 | 無職 |
|----|----|-------|----|----|
| 男性 | x | x | x | x |
| 女性 | ○ | x | x | x |

(B)

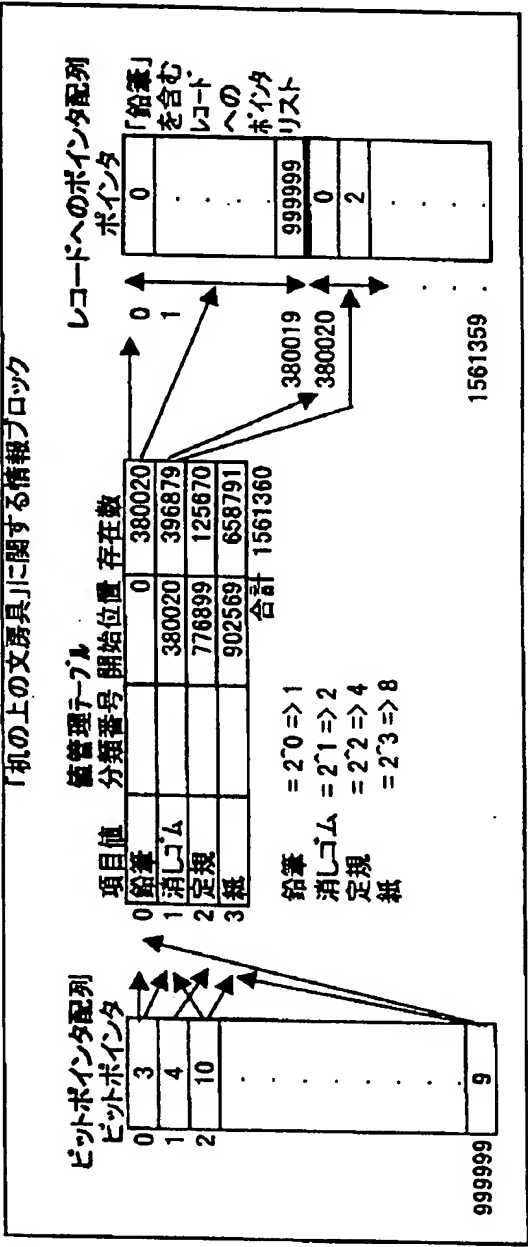
【図24】

マルチアンサ項目の説明図

| | |
|--------|---------|
| 0 | 鉛筆:消しゴム |
| 1 | 定規 |
| 2 | 紙:消しゴム |
| | |
| 999999 | 紙:鉛筆 |

【図25】

本発明の第10の実施例によるマルチアンサ対応情報ブロックの説明図



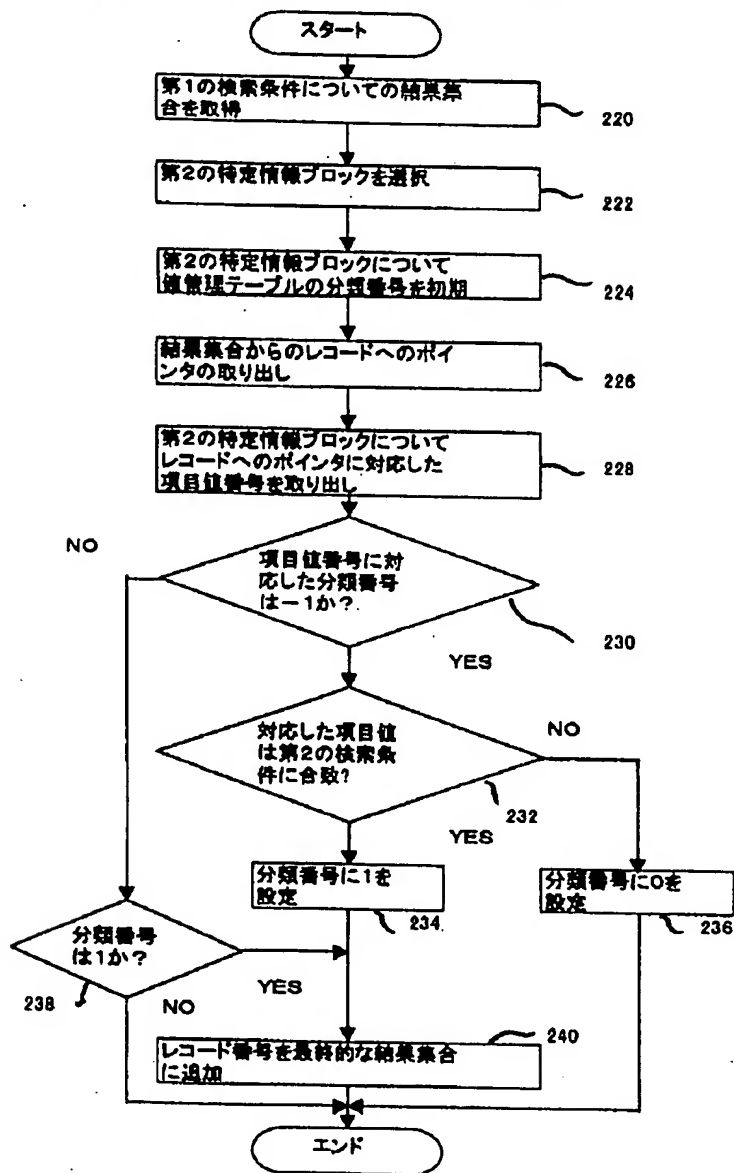
【図26】

本発明の第11の実施例による特殊値の取扱方法の説明図

| | 学生 | プログラマ | 教員 | その他 | フランク |
|------|----|-------|----|-----|------|
| 男性 | 81 | 5 | 0 | 5 | 6 |
| 女性 | 54 | 8 | 56 | 6 | 8 |
| フランク | 3 | 2 | 2 | 5 | 12 |

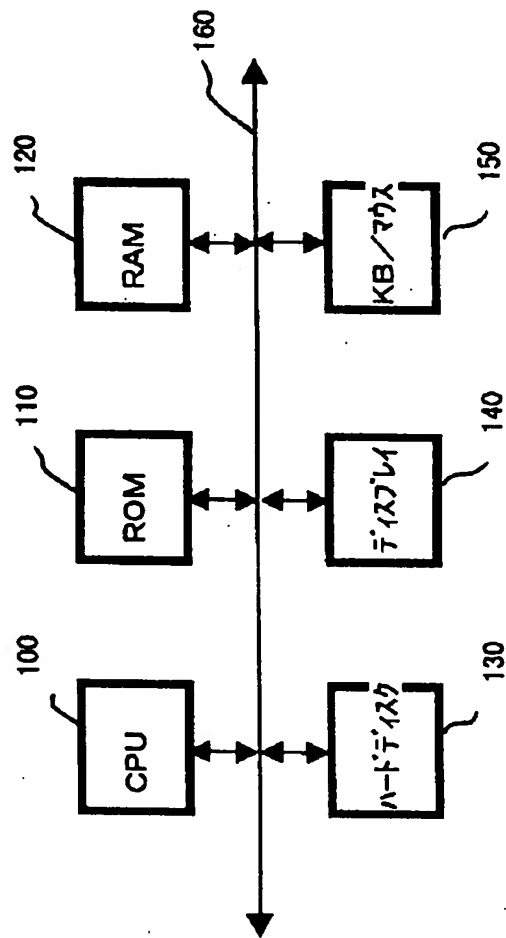
【図27】

本発明の第12の実施例による複数項目検索方法の動作フローチャート



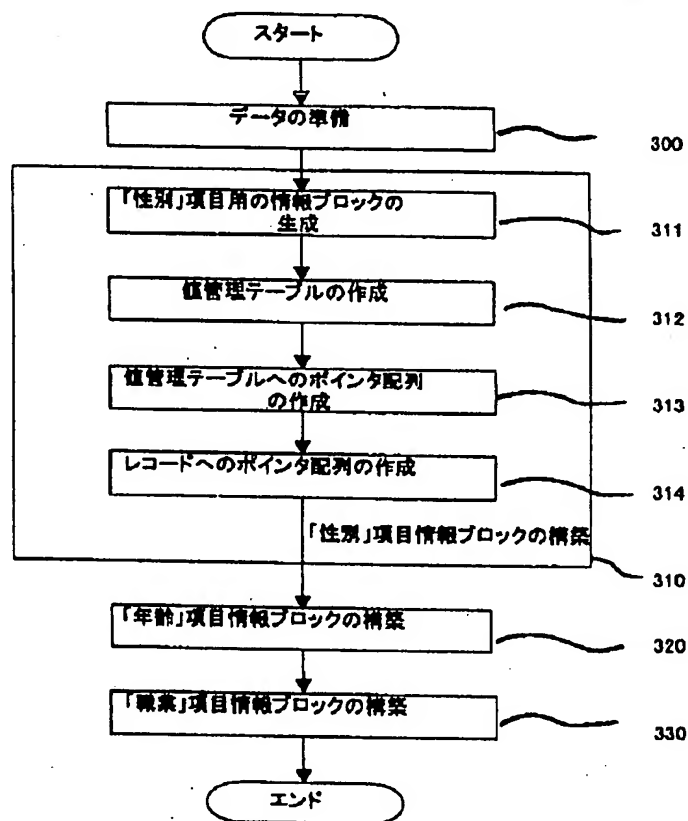
【図28】

本発明の一実施例による表形式データの検索・集計装置の構成図



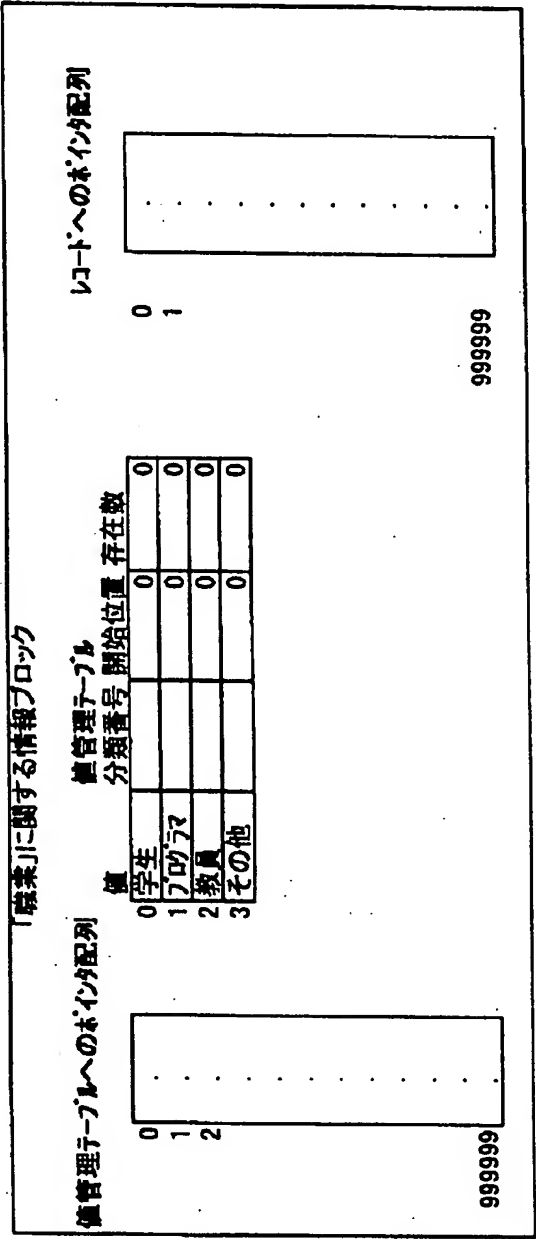
【図29】

表形式データに対する情報ブロックを構築する方法の説明図



【図30】

データ投入の準備と初期化の説明図

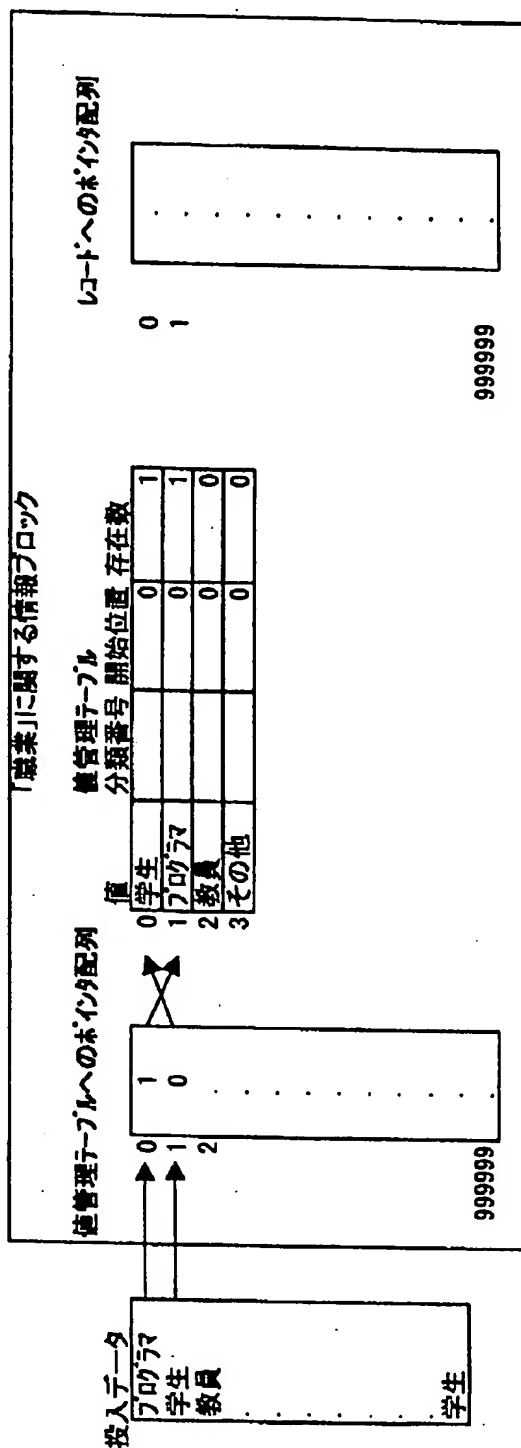


投入データ

| |
|-------|
| プログラマ |
| 学生 |
| 教員 |
| |
| 学生 |

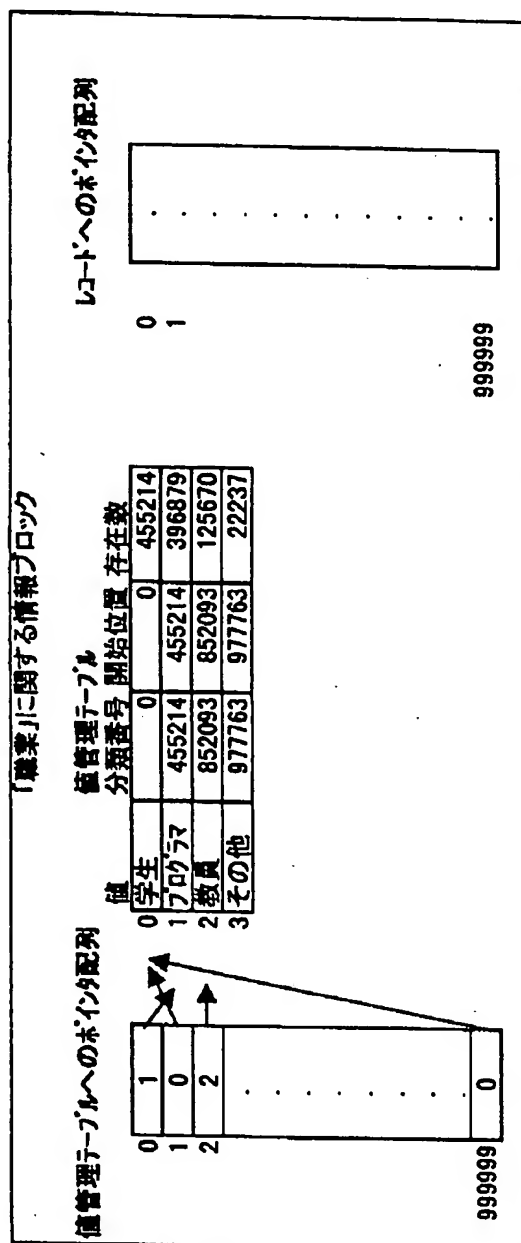
【図31】

データ投入の第1のパスの説明図



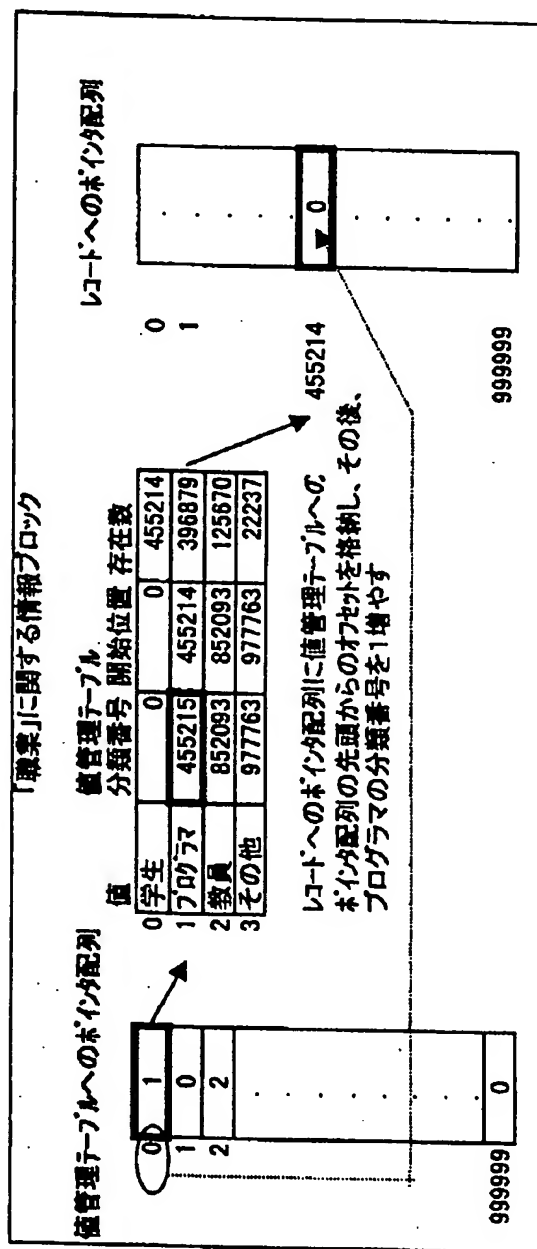
【図32】

データ投入の第2のパスの説明図



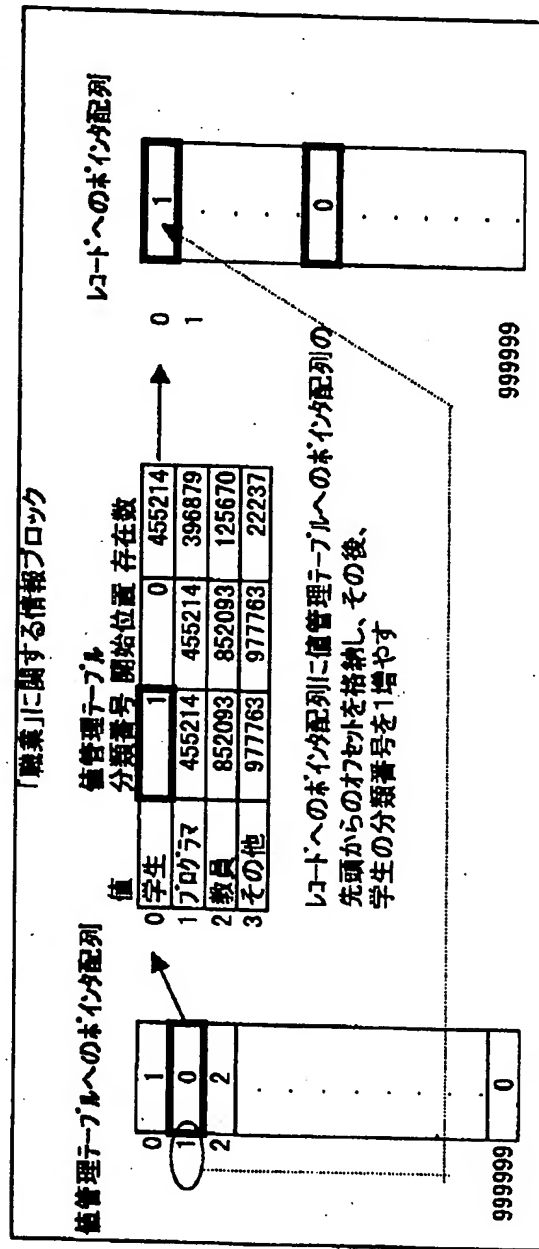
【図 3 3】

データ投入の第3のバスの説明図



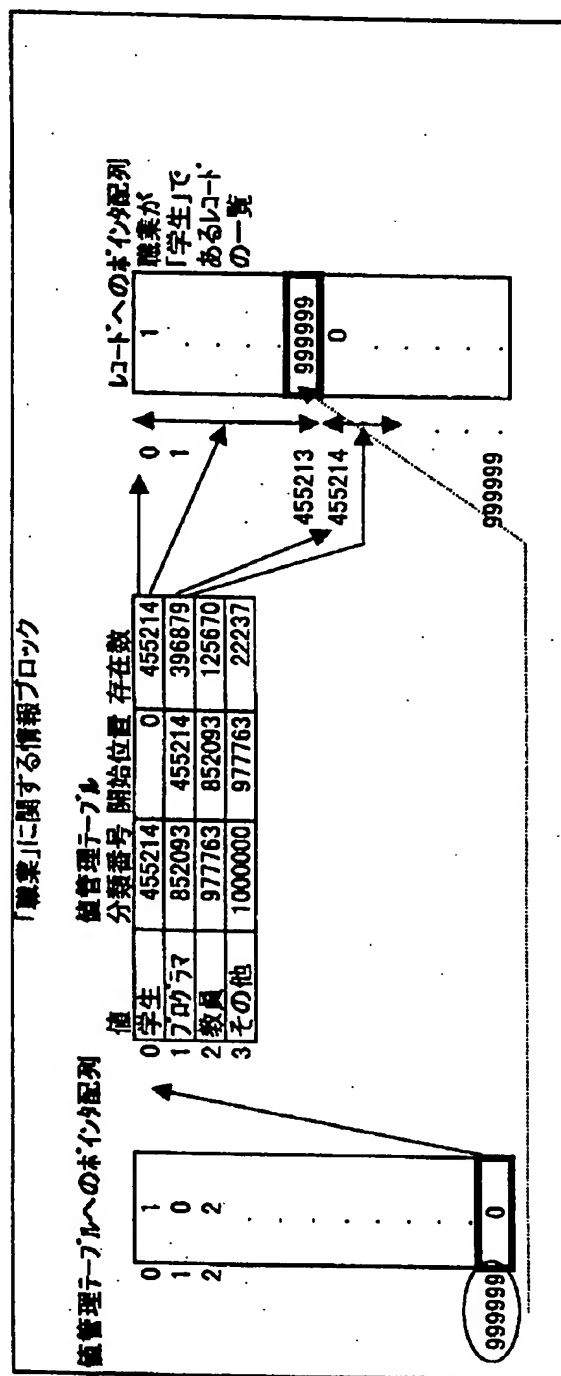
【図34】

データ投入の第3のバスの説明図



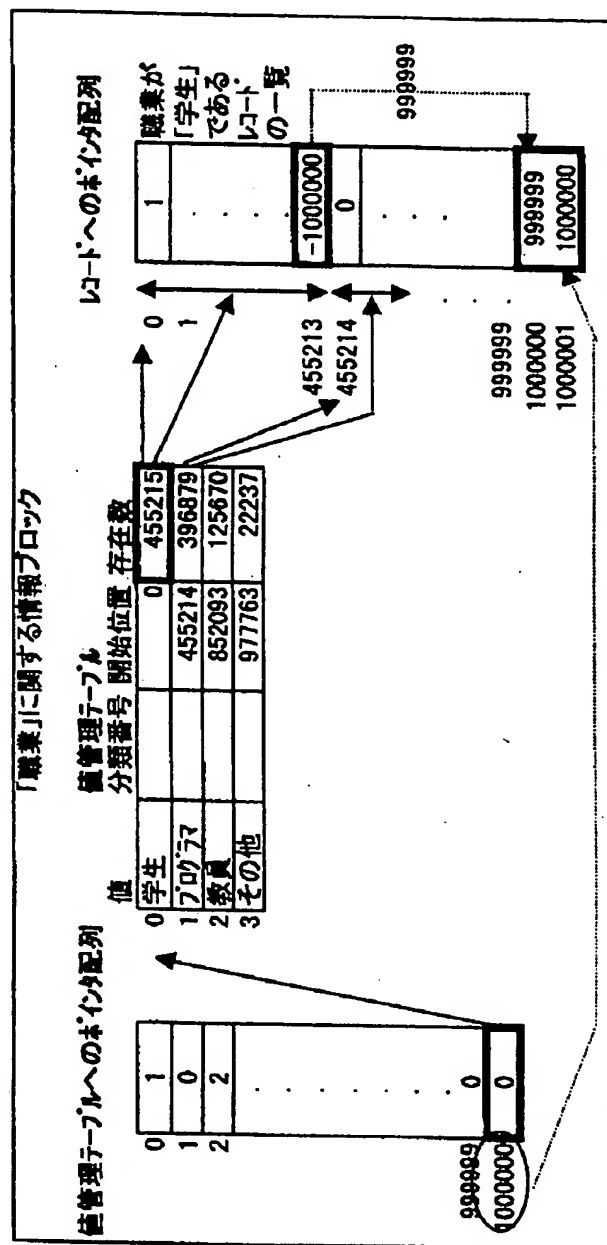
【図35】

データ投入の第3のパスの説明図



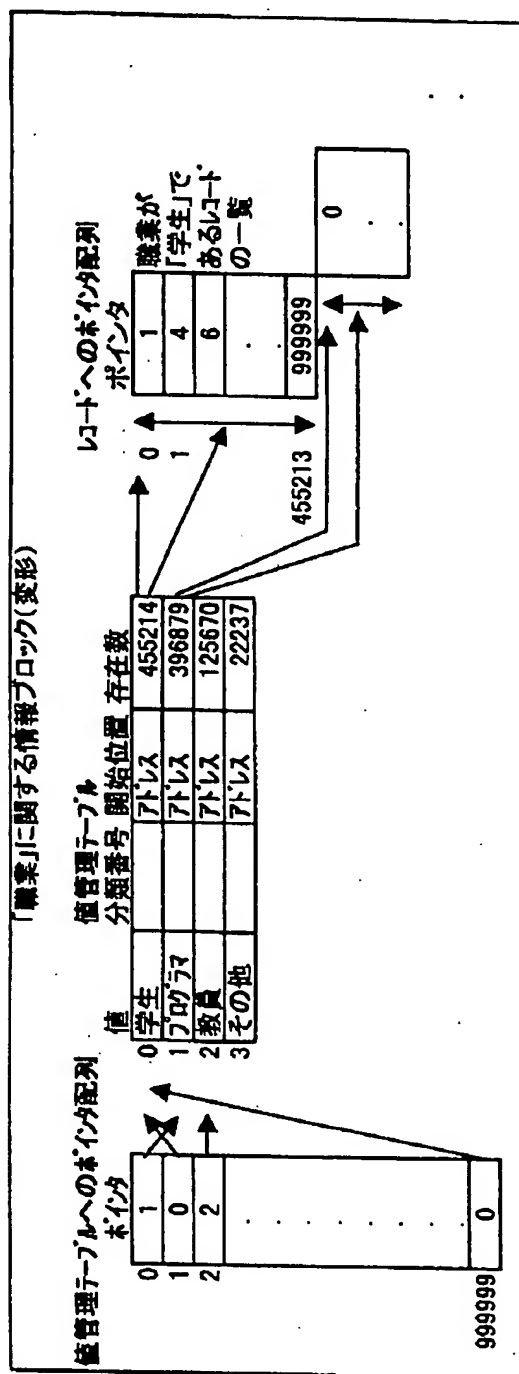
【図 3 6】

情報ブロックに対するデータ追加の説明図



【図37】

本発明の他の実施例による情報ブロックの説明図



【図38】

100万件のテスト用データの一覧

| レポート番号 | x1万 | x100 | x1 |
|--------|-----|------|----|
| 000000 | 00 | 00 | 00 |
| 000001 | 00 | 00 | 01 |
| . | . | . | . |
| . | . | . | . |
| 000100 | 00 | 01 | 00 |
| 000101 | 00 | 01 | 01 |
| . | . | . | . |
| . | . | . | . |
| 010000 | 01 | 00 | 00 |
| 010001 | 01 | 00 | 01 |
| . | . | . | . |
| . | . | . | . |
| 999999 | 99 | 99 | 99 |

【図39】

100万件の検索・集計のテスト結果

| ヒット件数 | ヒットフラグ | ポイント配列 | |
|---------|--------|--------|-------|
| | 検索 | 検索 | クロス集計 |
| 1000000 | 324 | 230 | 168 |
| 500000 | 280 | 185 | 83 |
| 250000 | 210 | 184 | 40 |
| 100000 | 205 | 127 | 15 |
| 10000 | 120 | 106 | 3 |
| 1000 | 35 | 95 | 0 |

【書類名】 要約書

【要約】

【課題】 本発明の目的は、従来のデータ表の機能を有すると共に大量のデータの検索・集計のスピードが大幅に向上した表形式データの検索・集計方法及び装置を提供することである。

【解決手段】 本発明による情報の項目に対応した項目値を含むレコードの配列として表される表形式データの検索・集計方法は、特定の項目に属する項目値に対応した項目値番号の順番に項目値が格納されている値管理テーブルと、レコードの順番に項目値番号を指定する情報が格納された項目値番号指定情報配列とを記憶装置に保持し、項目値番号指定情報配列から特定のレコードに対応した項目値番号を取得し、値管理テーブルに格納された項目値の中から、取得された項目値番号に対応した項目値を得る。

【選択図】 図5

【書類名】 出願審査請求書

【提出日】 平成10年8月11日

【あて先】 特許庁長官 伊佐山 建志 殿

【出願の表示】

 【出願日】 平成10年8月11日提出の特許願

 【整理番号】 ASIST9801

【請求項の数】 12

【請求人】

 【出願人との関係】 本人

 【住所又は居所】 神奈川県横浜市神奈川区松見町4丁目1101番地7
コートハウス菊名804号

 【氏名又は名称】 古庄 晋二

【代理人】

 【識別番号】 100070150

 【弁理士】

 【氏名又は名称】 伊東 忠彦

【手数料の表示】

 【予納台帳番号】 002989

 【納付金額】 116700